ENGINEERING

USCIPI Report 690

(1)

UNIVERSITY OF SOUTHERN CALIFORNIA

ESTIMATION AND DETECTION OF IMAGES DEGRADED BY FILM-GRAIN NOISE

by

Firouz Naderi

September 1976

Image Processing Institute

University of Southern California

University Park

Los Angeles, California 90007

Sponsored by

Advanced Research Projects Agency

Contract No. F-33615-76-C-1203

ARPA Order No. 3119

IPI

IMAGE PROCESSING INSTITUTE

USCIPI Report 690

ESTIMATION AND DETECTION OF IMAGES DEGRADED BY FILM-GRAIN NOISE

by

Firouz Naderi

September 1976

Image Processing Institute
University of Southern California
University Park
Los Angeles, California 90007

This research was supported by the Advanced Research Project Agency of the Department of Defense and was monitored by the Air Force Avionics Laboratory under Contract No. F-33615-76-C-1203, ARPA Order No. 3119.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Image Processing Institute
University of Southern California, University Park
Los Angeles, California 90007

2a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

2b. GROUP

3. REPORT TITLE
ESTIMATION AND DETECTION OF IMAGES DEGRADED BY FILM-GRAIN NOISE

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Technical Report, September 1976

5. AUTHOR(S) (First name, middle initial, last name)
Firouz Naderi

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
1 September 1976	160	35

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
F-33615-76-C-1203	
b. PROJECT NO.	USCIPI Report 690
c. ARPA Order No. 3119	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT
PUBLIC
Approved for release: distribution unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Advanced Research Projects Agency 1400 Wilson Boulevard Arlington, Virginia 22209

13. ABSTRACT

Film-grain noise is a term describing the intrinsic noise produced by a photographic emulsion during the process of recording an image on film. Although film-grain noise has been recently considered within the field of image processing, the nature of the noise is still somewhat misunderstood.

One goal of this study has been to use the theoretical and experimental results on film characteristics obtained by photographic scientists in order to define film-grain noise within the context of estimation theory. A detailed model for the overall photographic imaging system is presented. There are linear blurring effects at the initial and the final segments of this model to account for various optical and chemical degradations. The middle segment of the model represents signal dependence effects of film-grain noise and includes a nonlinear noise term. The accuracy of this model is tested by simulating images according to it and comparing the results to images of similar targets that were actually recorded on film. These comparisons point out that the model is a good representation of the photographic imaging system.

The restoration of images degraded by film-grain noise is considered in two different contexts - estimation theory and detection theory. Under the topic of estimation, a discrete Wiener filter is developed which explicitly allows for the signal-dependence of the noise. The filter adaptively alters its characteristics based on nonstationary first order statistics of an image. This filter is shown to

DD FORM 1 NOV 68 1473

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

have an advantage over the conventional Wiener filter.

In the case of extremely low contrast images digitized by a very small aperture, film-grain noise is so severe that conventional statistical resotration techniques have little effect. For use in this situation a heuristic algorithm is developed which incorporates some of the vision properties of the human observer. Bayesian detection theory is used to justify the procedure and to provide some insight into its use. This algorithm also explicitly includes the signal dependence of the noise and has the capability of greatly outperforming the human observer in locating objects corrupted by very severe noise.

Experimental results for modeling, the adaptive estimation filter and the Bayesian detection algorithm are presented.

* *

14. Key Words: Film-Grain Noise, Image Processing, Wiener Estimation, Bayesian Detection.

TO MY BROTHER JOHN.

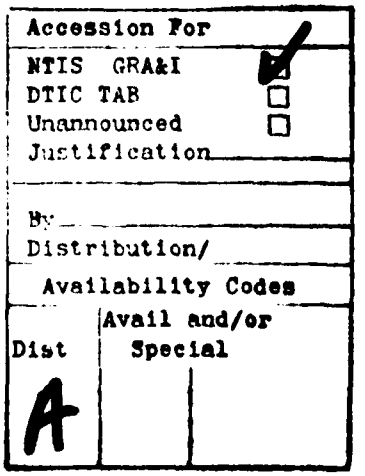
ACKNOWLEDGEMENT

I feel deeply indebted to the Chairman of my Dissertation Committee, Dr. Alexander Sawchuk. His advice and constant encouragement during the course of this research was truly invaluable.

I would also like to thank Professor Nasser Nahi for his guidance during my entire graduate study, and Professor John Pierce for his interest in this work.

I am very grateful to Dr. William Pratt under whose leadership the Image Processing Institute of USC grew and gave me and many of my fellow graduate students a chance to do research in this field.

My thanks also goes to Dr. Hans Zweig for his many helpful suggestions during preparation of Chapters 3 and 4.



ABSTRACT

Film-grain noise is a term describing the intrinsic noise produced by a photographic emulsion during the process of recording an image on film. Although film-grain noise has been recently considered within the field of image processing, the nature of the noise is still somewhat misunderstood.

One goal of this study has been to use the theoretical and experimental results on film characteristics obtained by photographic scientists in order to define film-grain noise within the context of estimation theory. A detailed model for the overall photographic imaging system is presented. There are linear blurring effects at the initial and the final segments of this model to account for various optical and chemical degradations. The middle segment of the model represents signal dependence effects of film-grain noise and includes a nonlinear noise term. The accuracy of this model is tested by simulating images according to it and comparing the results to images of similar targets that were actually recorded on film. These comparisons point out that the model is a good representation of the photographic imaging system.

The restoration of images degraded by film-grain noise is considered in two different contexts - estimation theory and

detection theory. Under the topic of estimation, a discrete Wiener filter is developed which explicitly allows for the signal-dependence of the noise. The filter adaptively alters its characteristics based on nonstationary first order statistics of an image. This filter is shown to have an advantage over the conventional Wiener filter.

In the case of extremely low contrast images digitized by a very small aperture, film-grain noise is so severe that conventional statistical restoration techniques have little effect. For use in this situation a heuristic algorithm is developed which incorporates some of the vision properties of the human observer. Bayesian detection theory is used to justify the procedure and to provide some insight into its use. This algorithm also explicitly includes the signal dependence of the noise and has the capability of greatly outperforming the human observer in locating objects corrupted by very severe noise.

Experimental results for modeling, the adaptive estimation filter and the Bayesian detection algorithm are presented.

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont'd)

TABLE OF CONTENTS (Cont'd)

LIST OF FIGURES

LIST OF FIGURES (Cont'd)

LIST OF FIGURES (CONT'D)

LIST OF FIGURES (CONT'D)

LIST OF TABLES

Chapter 1

INTRODUCTION

Image restoration and enhancement are terms used to describe methods of removing different types of degradations in imaging systems. With the advent and improvement of digital computers, investigators have concentrated for the past decade on digital techniques to deal with the large quantities of pictorial data in image processing.

The degradations that an imaging system imposes on a picture are often of two types. There are degradations that can be described by smoothing operations, including effects due to the finite resolution of sensors and various optical and chemical effects. There are also degradations due to measurement errors, and these are usually referred to as noise. Film-grain noise is one such disturbance.

The goal of this study is to unite some results of photographic scientists in modeling film-grain noise with some recent developments in the field of digital image processing to improve the quality of images degraded by severe film-grain noise. In order to make this study useful to workers both in the fields of digital image processing and photographic science, a summary of relevant past results is presented for each topic.

The material presented here is divided into three major topics: modeling, estimation and detection. The following is an outline of this study and a summary of the research contributions in each. In Chapter 1 we examine the problem of film-grain noise by considering the physical structure of photographic film. In particular we discuss various factors effective film-grain noise during the exposure and development.

The mathematical modeling of film-grain noise is the subject of Chapter 2. First we review some simple models suggested by different researchers. We then use published experimental results on film characteristics to investigate the functional form of the standard deviation of the noise. The trade-off between linear modeling with signal dependent noise and nonlinear modeling with signal-independent noise is also discussed in this chapter. A subjective evaluation procedure for various models is suggested and the chapter concludes with a detailed model flexible enough to accomodate different degradations imposed by an imaging system.

Chapter 3 discusses restoration of film-grain noise degraded images under the topic of estimation. A discrete linear minimum mean square error filter is developed which explicitly allows for the signal-dependence of the noise. This filter adaptively alters its characteristics based on the non-stationary first order statistics of

the image. This filter is shown to have an advantage over the conventional Wiener filter.

Chapters 4 and 5 consider the detection of extremely low contrast images which have been almost completely obliterated by film-grain noise. In Chapter 4 we provide mathematical justification for an algorithm called "noise cheating" previously used by other researchers. In Chapter 5 we first improve on "noise cheating" to make it applicable to a broader class of images and to utilize some of the signal-dependent effects from our model. We also develop a more versatile Bayesian detection scheme which yields superior results and extends the usefulness of the technique.

1.1 Structure of Photographic Film

Photographic film is probably the most commonly used device for detecting or recording light. Let us consider a cross section of a typical commercial film for use in ordinary cameras. As shown in the schematic drawing of Fig. 1.1-1, film consists of an emulsion supported on a firm base. The emulsion is composed of a clear transparent gelatin containing a suspension of silver halide particles. These particles, or film grains as they are called, are photo sensitive and typically consist of a lattice of silver ions and halide ions. Upon exposure to light, the state of some of these grains will change in such a manner as to record the spatial distribution of the

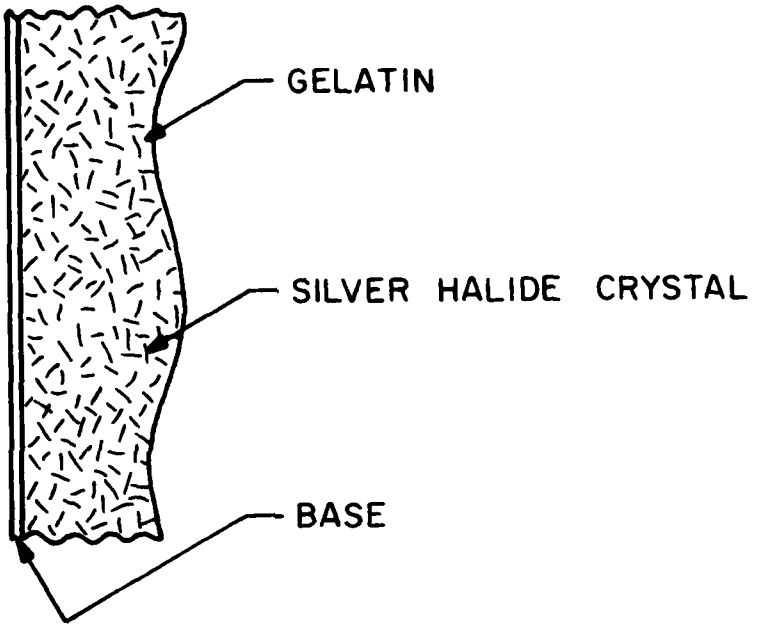
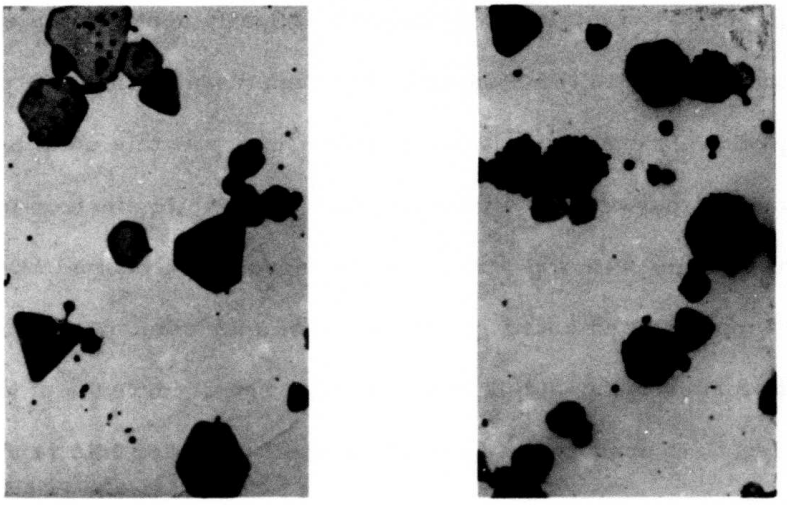


Figure 1.1-1. Schematic drawing of a cross section of a photographic film.

exposing light. The process of image formation basically involves two separate steps:

1. Formation of Latent Image. When exposed to light, photons of sufficient energy will be absorbed by the silver halide, thus raising an electron from the valence band to the conduction band. Assuming the halide crystal was initially electrically neutral, the elevation of the electron to the conduction band must simultaneously produce a positively charge entity--a positive hole. At this point one of two events may occur. Either the free and mobile electron may recombine with a positive hole to reproduce a neutral crystal plus some released energy or the electron may combine with a silver ion to form a quasi-stable silver atom. The occurrence of these two events is probabilistic and as we shall see later this randomness is indeed one of the causes of film-grain noise.

If a sufficient number of photons are absorbed by a grain and they in turn result in silver atoms, these atoms form tiny silver specks called development centers on the surface of the exposed grain as shown in Fig. 1.1-2. The size of these specks are larger on the grains that were heavily exposed and much smaller on the grains that did not absorb enough photons. Naturally there are crystals that are completely left unchanged because they were not exposed or were exposed very little. Photographic scientists use the term latent image to describe the film at this point. Although no visible image



(a)
Low Intensity

(b)
High Intensity

Figure 1.1-2. Electron micrograph for a uniform exposure showing development centers consisting of aggrigates of silver atoms on silver bromide emulsion grains. [2]

has yet been formed, the state of some of the halide grains has been changed as to retain the spatial distribution of the exposing light. These grains with development centers will render a visible image in the second step of image formation process.

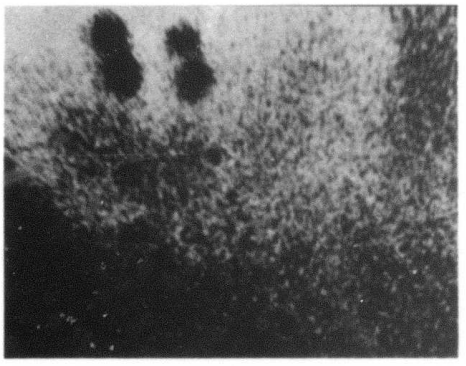
2. Development of Latent Image. To produce a visible image, the exposed film is immersed in a developer solution. The developer acts as a reducing agent and completely changes the halide crystals with stable development centers to opaque metallic silver grains. The rate or the probability that they are converted, however, depends on the size and the number of these silver specks. It is believed that if three to six silver atoms or more are formed during the exposure, then a stable development center is formed. Undeveloped crystals are washed away into the developer solution leaving clear transparent gelatin in their place.

The question that arises naturally is why photographs appear continuous when the congregations of metallic silver grains are scattered randomly within a clear gelatin. The answer clearly is that photographs do indeed appear as random distribution of opaque grains within a transparent gelatin if they are viewed under sufficient magnifications as shown in Fig. 1.1-3. However since the human visual system has a limited resolving capability and film grains are small in size, (see Table 1.1-1), regions in the film consisting of a

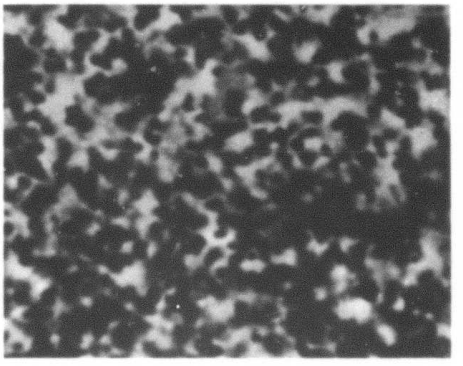
Continuous tone positive proof

The negative record

A 16 x enlargement



A 100 x enlargement



A 1000 x enlargement

Figure 1.1-3. A series of pictures demonstrating the grainy appearance of photographs under sufficient enlargement [1].

considerable number of opaque metallic silver grains appear to be a homogeneous shade of black. On the other hand, areas where there are no grains or where the grains are very sparse appear to be transparent.

As discussed earlier in this section, film is a device for recording light. Indeed, when one looks at a developed photograph, one gets a subjective feeling for the spatial distribution of the exposing light. Black regions on the photograph, for example, indicate a high intensity of the exposing light. In order to establish some quantitative measure of the exposing light or some parameter such as its intensity let us view the film as a system and define an input, output and a transfer function.

Plate or Film	Average Diam. (μm.)	Average Area $(\mu m.)^2$
High-resolution film	0.048	0.00188
Motion-picture-positive film	.30	.07
Positive-type film	.63	.31
Fine-grain roll film	.79	.49
Portrait film	.88	.61
High-speed roll film	1.09	.93
X-ray film	1.71	2.30

Table 1.1-1. Grain size for various type of films. [2]

1.2 Film Viewed as a System

The input to the film is the incident light, or more generally, the exposure E defined as

$$E = I \cdot t \qquad (1.2\text{-}1)$$

where I is the intensity of the incident light and t is the exposure time. The most common quantitative measures of the output of a photographic film are either optical density or transmittance. For a region R in a developed film, transmittance is measured in the following manner: the region is illuminated with a light of constant intensity I_i and then the intensity of the light transmitted to the other side of the film through this region is measured. Denoting this intensity by I_t, the transmittance value for this region is defined as

$$T = \frac{I_t}{I_i} \qquad (1.2\text{-}2)$$

and optical density is defined as

$$D = -\log_{10} T = \log_{10} \frac{I_i}{I_t} \quad . \qquad (1.2\text{-}3)$$

When we look at a photograph, we perceive a spatial variation in incident light intensity. Photographic film converts this variation by appropriate exposure and development to spatial variations in optical density or transmittance of the film. From eqs. (1.2-2) and (1.2-3) we note that since $I_i \geq I_t$, optical density is a nonnegative quantity and transmittance T varies between 0.0 and 1.0. As we shall see in Chapter 2, the measured optical density for a region is directly proportional to the mass of metallic silver deposited in that region during the process of exposure and development.

Suppose the incident light exposing a particular region R has a high intensity. Most of the halide grains within this region will be transformed to metallic silver grains. Now, since silver grains are opaque, in the process of measuring the optical density for this region the transmitted intensity I_t of eq. (1.2-3) will be much less than intensity I_i. Therefore as shown by eq. (1.2-3), a large value for optical density will be recorded. Thus high optical density for a region corresponds to a large mass of silver in that region which in turn is indicative of high intensity of the exposing light.

So far in this section we have established quantitative input and output for a film. What remains to be discussed is a transfer function linking the input and the output together. Around 1900, two photographic scientists, Hurter and Driffield, performed a series of experiments. They discovered that in macroscopic regions, the logarithm of the exposure E (input to the photographic system) and the optical density (output of the photographic system) are related through a function which has become known as the Hurter and Driffield (H and D) curve or D-log E characteristic curve. A typical H and D curve is shown in Fig. 1.2-1.

If the film is operated in the monotonic region of this curve, a measured optical density would uniquely determine the input exposure, at least in the macroscopic regions for which this curve is obtained.

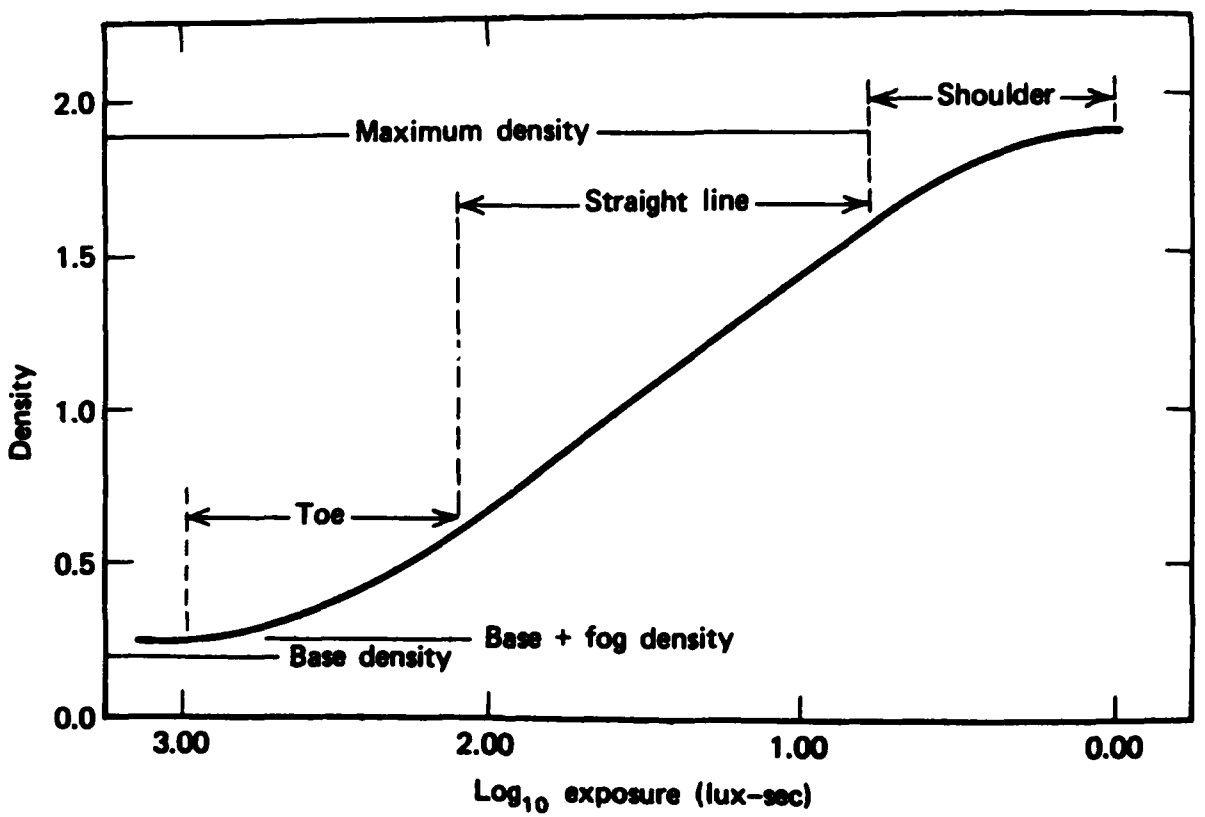


Figure 1.2-1. A typical Hand D curve for a medium speed film [2].

Thus by measuring optical density in various regions of a developed film one can determine the spatial distribution of the exposing light.

However, if we measure the optical density of a developed film not in macroscopic regions but in microscopic regions, we note that the measured optical density does not relate to the input exposure in a straightforward manner and film behaves more like a stochastic system. This is what we will examine in the next section.

1.3 Film-Grain Noise

Suppose we expose a film with light having absolutely no spatial variation, or in other words, perfectly constant intensity. If this film is viewed with a naked eye after careful development we see that a constant shade of gray has been registered across the film. So far everything is according to expectations; an input with a constant level has resulted in a subjectively constant output. If we measure optical density in various macroscopic regions, perhaps two or three millimeters in diameter, as predicted by the H and D curve, the same value for the optical density is measured in the various regions.

However, if we use a microdensitometer with a small aperture of only a few microns in diameter to measure optical density for several microscopic regions in the film, we note that different readings are obtained for different regions even though the exposing light for all of the regions was the same. Considering photographic film as a space-invariant system, these fluctuations in the output

from one region to another, when in fact the input in all of the regions is the same, indicates that the system is not noise free. There are several reasons for this phenomenon, some of which are discussed below:

1. As discussed in section 1.1, photographic film is composed of microscopic silver grains and thus is inherently a discrete medium. It should be expected, therefore, that under sufficient magnification it would appear inhomogeneous. In fact, if transmittance T defined in eq. (1.2-2) could be measured at any point (x,y) it would be either zero or one depending on whether the point was covered by an opaque silver grain(s) or not. Therefore film can be taken to be a binary medium on a microscopic scale.

2. Even though the incident light may be spatially uniform over the extent of the film, from a quantum viewpoint the number of photons arriving at the surface of the film per unit time is not constant over the entire surface but has a poisson distrubution. Thus, each silver halide grain in the film is struck with a different number of photons and therefore some have more chance to reduce to metallic silver during the development. This effect is particularly significant when the incident light is of a very low intensity.

3. As discussed in section 1.1, both the formation of the latent image and its development are probabilistic in nature. Two seemingly identical silver halide grains struck with the same number of

photons react differently. One might eventually be converted to a silver grain whereas the other might not.

4. It has been established that silver grains for a uniformly exposed and developed film are randomly scattered throughout the film [3]. Their distribution is nearly binomial and can be approximated by poisson if the grains are assumed indefinitely small. If a microdensitometer is used to scan the film, the number of grains that fall within the area covered by the scanning aperture is different at different locations. Since measured optical density for a region is directly proportional to the mass of metallic silver deposited in that region, the randomness associated with the number of grains that fall within various regions contributes to randomness in measured optical density.

5. It is practically impossible to make an emulsion with constant size grains. Thus the variation in grain size will also affect measured optical density due to relationship between optical density and the mass of silver.

In summary, the fluctuations in the number, size and configuration of grains from one small subregion to the next causes corresponding fluctuations in the optical density (or transmittance) of microscopic regions in a uniformly exposed and developed film, and this produces the noise in the photographic system.

When this noise is measured objectively by various types of instrumentation, it is called granularity, whereas the subjective impression of film noise in a photographic enlargement is called graininess.

References

[1] J. F. Hamilton, W. H. Lawton and E. A. Trabka, "Some Spatial and Temporal Point Processing in Photographic Science," Invited paper at the International Conference on Stochastic Point Processes: Statistical Analysis, Theory and Application, IBM Research Center, Yorktown Heights, NY, August 2-7, 1971.

[2] W. Thomas, SPSE Handbook of Photographic Science and Engineering, John Wiley & Sons, NY, p. 768, 1973.

[3] C. E. K. Mees and T. H. James, The Theory of the Photographic Process, 3rd Edition, Macmillan, NY, 1966, p. 525-527.

Chapter 2

MODELING

One of the aims of science is to understand the nature of the real world based on simplified artificial models. The laws and relationships that evolve from these idealized models are approximations. In the case of most real world processes, these idealized models must be altered or supplemented to reflect the complex nature of the process. The tradeoff is usually between the utility of the simplified models and the accuracy of the more complex ones.

As briefly reviewed in the first chapter, the formation of an image on photographic film is a highly complex optical and chemical process. Modeling this process with a high degree of accuracy, if at all possible, often results in models that are too complex and unsuitable for use in any subsequent mathematical processing. On the other hand, the reliability of restoration techniques such as estimation or detection is directly related to the degree by which the underlying mathematical model simulates the actual physical process. Therefore, oversimplification of the model makes the subsequent restoration technique very sub-optimal.

With these thoughts in mind, the goal of this chapter is to present a detailed model for the overall imaging system which is

reasonably accurate and yet mathematically tractable.

In section 2.1 we review some simplified models which give us a general expression for the standard deviation of noise of the photographic system. In section 2.2 we see how various quantities, such as the standard deviation, can be used as an objective measure of graininess of films. Some experimental results are analyzed in section 2.3 and this analysis leads to a more accurate form for the standard deviation. The probability distribution of optical density is the subject of section 2.4. Using the results of the first four sections, in section 2.5 we obtain some models for the optical density in the context of estimation theory. In section 2.6 we discuss the tradeoff between various models. A procedure for the evaluation of various mathematical models of the photographic process is outlined in section 2.7, and a detailed model for the imaging system is suggested in section 2.8.

2.1 Standard Deviation of Film-Grain Noise

In 1913 Nutting gave a simplified model for a developed photographic layer, expression optical density of the image in terms of the number and size of developed grains [1]. Suppose a film is exposed to a spatially uniform light and then developed. Nutting assumed that the developed emulsion layer may be divided into k elementary levels each about one grain thick as shown in Fig. 2.1-1. The grains are

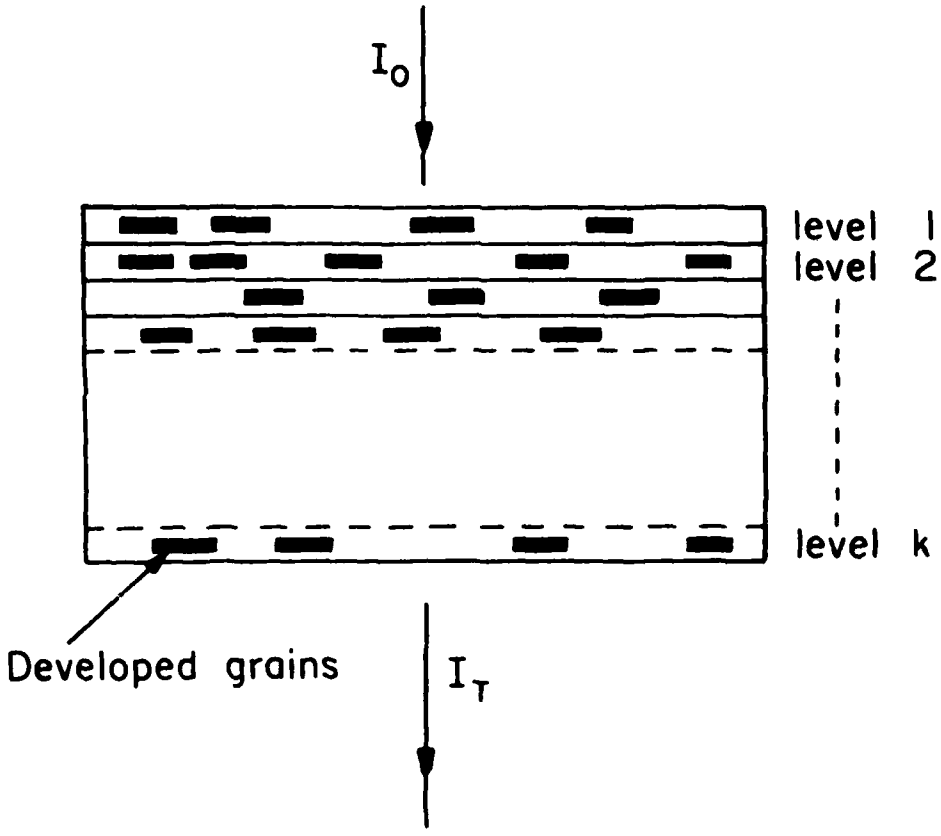


Figure 2.1-1. A simplified model of the developed photographic emulsion used by Nutting.

assumed randomly distributed within each level with a negligible degree of overlap. For the above model Nutting showed [2] that the optical density D can be written as

$$D = .43 \frac{na}{A} \qquad (2.1\text{-}1)$$

where a is the size of the developed grains (which is taken to be the same for all the grains), A is the size of the scanning aperture used to measure the optical density D and n is the total number of grains that fall within the aperture area at any given point.

Despite the many simplifying assumptions that were made in deriving eq. (2.1-1), Nutting's model has proved of lasting benefit and this equation and variations of it are still widely used.

As mentioned in chapter 1, the grains in a uniformly exposed and developed film are randomly distributed within the gelatin. Therefore n in eq. (2.1-1) is a random variable. This randomness in turn gives rise to fluctuations in the measured optical density D of the uniformly exposed film. As we said earlier it is this fluctuation in the optical density of a uniformly exposed and developed film that accounts for the noisy nature of photographic film as a measuring device for light. The standard deviation of D may be taken as a measure of this noise and can be calculated from eq. (2.1-1). Letting E be the ensemble average operator we have:

$$E(D) = .43 \frac{E(n) \cdot a}{A} \qquad (2.1\text{-}2)$$

and

$$\sigma_D^2 = .43 \frac{\sigma_n^2 \cdot a}{A} \qquad (2.1\text{-}3)$$

where σ_D^2 and σ_n^2 are variances of D and n respectively. The distribution of n can be modeled by a binomial distribution [3] or, if $\frac{a}{A}$ is small, by a poisson distribution. Since the first two moments of the poisson distribution are equal, we have $E(n) = \sigma_n^2$, and using this fact in the above two equations we get

$$\sigma_D = \frac{\sqrt{.43a}}{\sqrt{A}} (\overline{D})^{\frac{1}{2}} \qquad (2.1\text{-}4)$$

where $\overline{D} = E(D)$. Equation (2.1-4) indicates that the fluctuation in optical density is dependent on three main factors:

a) The fluctuation is inversely proportional to the scanning aperture used in measuring D; it increases as the aperture decreases.

b) As seen in Table 1.1-1 the mean grain size varies for various films. Equation (2.1-4) points out that the fluctuation in the optical density is greater in high speed films which utilize large grains and is smaller in high resolution films which employ smaller grains.

c) The variation in density is dependent on the mean density $\overline{D}$. Thus if two frames of a film are exposed uniformly, one with

exposure E_1 and the other with exposure E_2, and these frames are then scanned with the same size aperture, the fluctuation in the density of the frame exposed to the higher exposure (which is thus developed to a higher mean density) is greater. Some typical values for σ_D are shown in Table 2.1-1.

As we will see in the next section, σ_D can be used as a measure of the graininess of a film.

2.2 Objective Measures of Graininess

Graininess is the impression of inhomogenity in a magnified image. It has been of primary importance to the photographic scientists to establish an objective measure to rank various types of films according to their subjective graininess.

Material	σ(D)
Tri-X Aerecon Type 8403	0.048
Plus-X Aerecon Type 8401	0.034
Panatomic-X Aerial Type 3400 (Estar thin base)	0.020
Kodachrome II for daylight	0.012
Special high-definition aerial film (gray base) Type SO-243	0.0074
High-resolution plate	0.0025

Obtained for recommended development and under the following scanning conditions: Net density of sample above support = 1.0; circular scanning-aperture diameter = 48 μm.

Table 2.1-1. Standard deviation of optical density for typical Kodak films and plates [4].

Bricout [5] was among the first to suggest exposing various types of films to a uniform exposure and then using $\sigma(D)$, described in eq. (2.1-4), as an objective measure of their graininess.

Trying to establish a single experimental quantity to measure the graininess of photographic materials, Selwyn [6] in 1935 published a paper in which he hypothesized that the probability $P(D, \Delta D)$ that an area A, chosen at random on a uniformly exposed film, would have a density different from the mean density of the film by an amount in the range of $[D, D+\Delta D]$, is a function of A, D, ΔD and a constant G which he called the "graininess" constant. He theorized that $P(D, \Delta D)$ is proportional to ΔD for vanishing by small values of ΔD so that

$$P(D, \Delta D) = f(G, D, A) \cdot \Delta D \qquad (2.2\text{-}1)$$

where f is a function to be determined. Following this hypothesis and using dimensional theory he showed the functional form of f to be:

$$P(D, \Delta D) = \frac{A}{\pi G^2} e^{-(D^2 A/G^2)} \cdot \Delta D \, . \qquad (2.2\text{-}2)$$

In the derivation of the above equation he assumed the size of the grains to be sufficiently small with respect to A so that adjacent nonoverlapping areas are independent.

Equation (2.2-2) suggests that the optical density has a Gaussian distribution with variance $\frac{G^2}{2A}$. The relationship

$$G_D = (2A)^{\frac{1}{2}} \cdot \sigma_D \qquad (2.2\text{-}3)$$

has become known as Selwyn's law and G is called the granularity or Selwyn coefficient. This coefficient was suggested as an objective measure of graininess. In eq. (2.2-3), the subscript on G is included to emphasize that the relationship defined through this equation makes sense only for a given mean density. As predicted by eq. (2.1-4) a change in the mean density results in a change in σ_D which in turn alters the value of the constant G.

2.3 A More Correct Version of the Standard Deviation

After Selwyn introduced the notion of granularity constant, many workers within the next twenty years performed a series of experiments in connection with Selwyn's law [5], [7], [8]. Their conclusions were quite varied with various experimenters reporting the exponent in the eq. (2.2-3) to be anywhere in the range of 0.265-0.5. However, due to the unreliability of the measuring devices, particularly the microdensitometers, most of the experimental results of this period are of questionable value.

Finally, Higgins and Stultz [9], armed with the results of previous researchers and improved measuring apparatus set out to experimentally examine Selwyn's law. Their experiment resulted in a set of data published in 1959. In this experiment they examined four types of films ranging from high speed Kodak Royal-X to fine-grained

Eastman Fine Grain Release Positive film.

Six samples of each of the four films were uniformly exposed and developed to six different optical density levels. Then each of the samples were scanned with seven different apertures ranging in size from 7.25 μm to 384 μm in diameter. The standard deviation σ_D was calculated for the data obtained with each of the seven apertures and values of G were then tabulated using eq. (2.2-3).

The results of Higgins and Stultz experiments are summarized in Figs. 2.3-1 to 2.3-4. In each figure, G is plotted against the square root of the scanning aperture area for all of the six optical density levels. From eq. (2.2-3) it is clear that Selwyn's law predicts that the graphs in these figures should be horizontal lines for each given density. Higgins and Stultz concluded that their data verified Selwyn's law and the deviation from horizontal lines were attributed to measuring errors and the impracticality of producing truly uniformly exposed films.

From Figs. 2.3-1 to 2.3-4 it is apparent that G is not a constant for a very wide range of apertures. The failure appears to be consistently pronounced at large aperture sizes. Furthermore, in the case of coarse-grained Royal-X film, the smallest aperture seems to be outside the range for which Selwyn's law holds. This may be expected because Selwyn assumes grain sizes to be much smaller than the scanning aperture in the deviation of eq. (2.2-2).

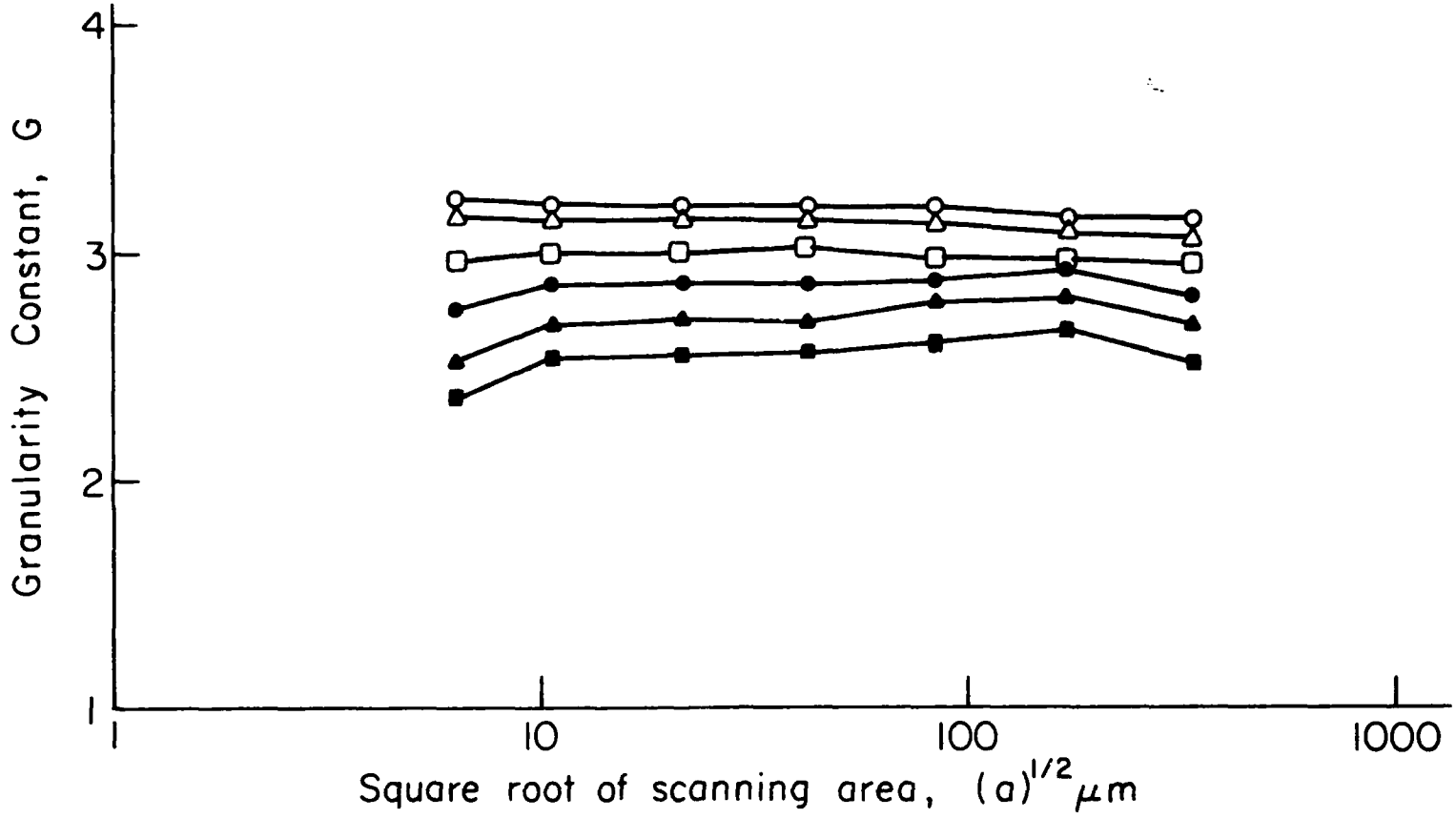


Figure 2.3-1. Plot of granularity constant G vs. square root of the scanning aperture area for Kodak Royal-X Pan Film [9]. Density levels: O 1.12; Δ 0.94; □ 0.71; ● 0.49, ▲ 0.31; ■ 0.20.

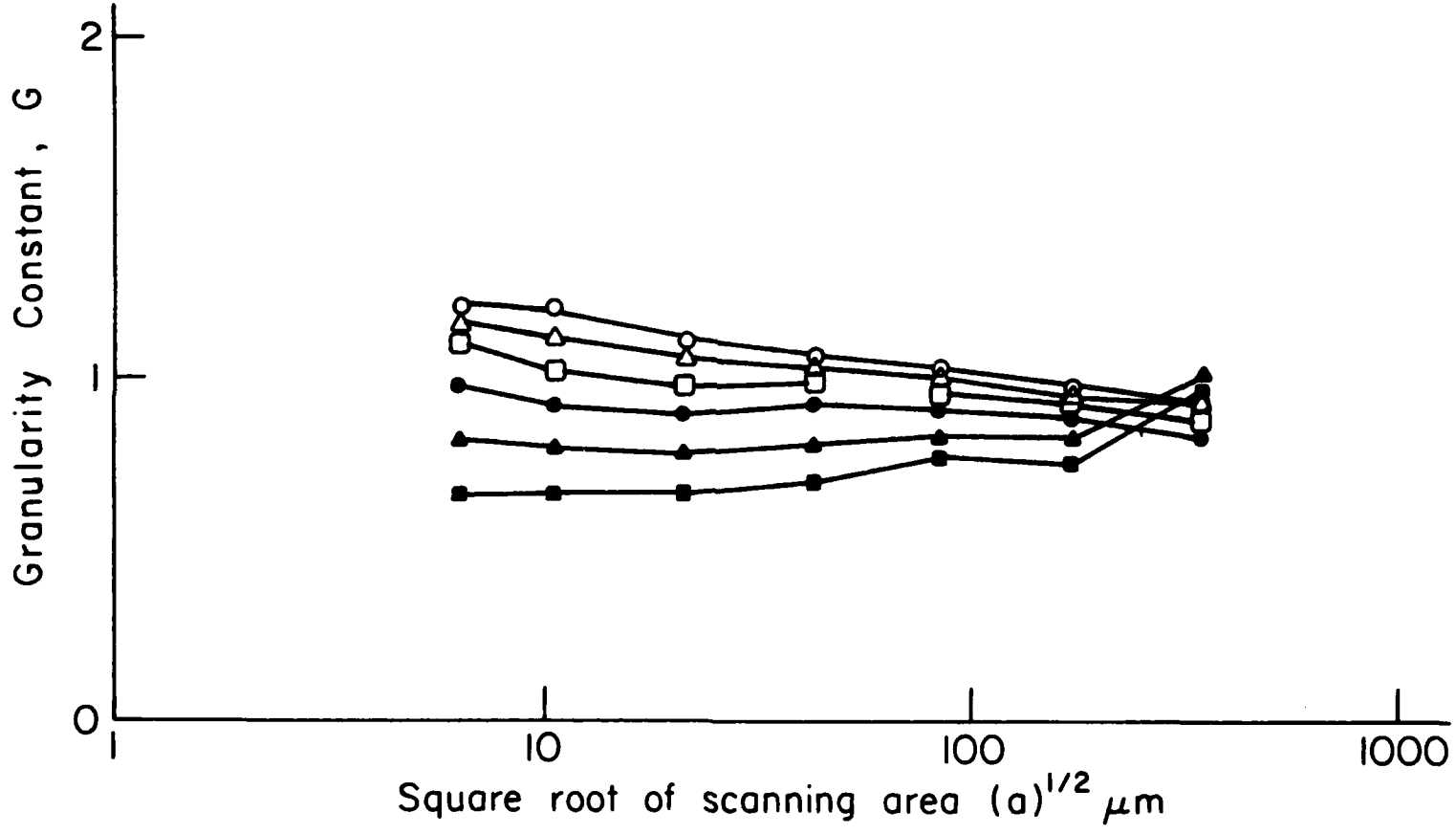


Figure 2.3-2. Plot of granularity constant G vs. square root of the scanning aperture area for Kodak Pan-X Film [9]. Density levels: O 0.96, Δ 0.75, □ 0.55, ● 0.39, ▲ 0.24; ■ 0.13.

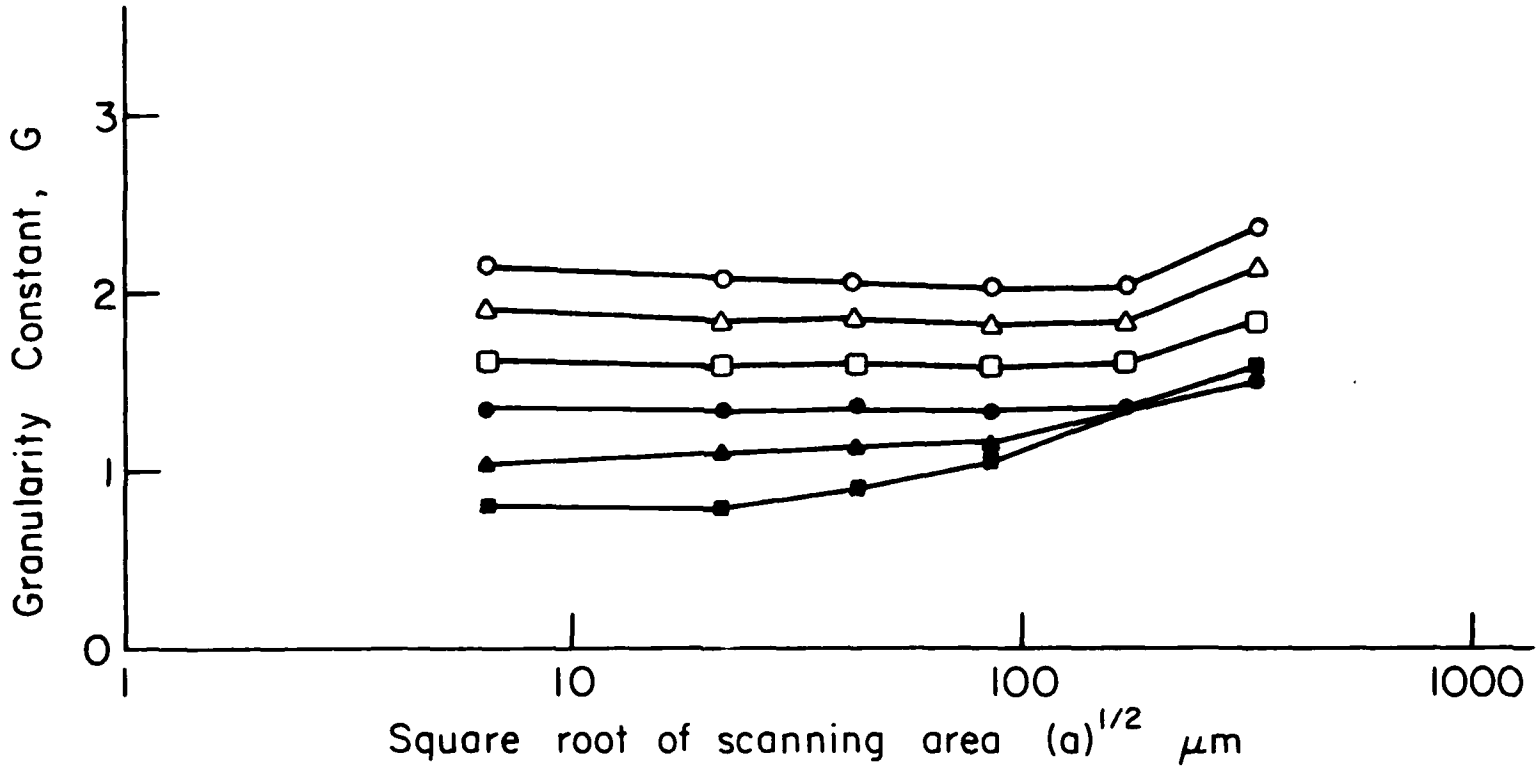


Figure 2.3-3. Plot of granularity constant G vs. square root of the scanning aperture area for Kodak Pan-X film [9]. Density levels: ○ 1.58; △ 1.21; □ 0.84; ● 0.54; ▲ 0.32; ■ 0.18.

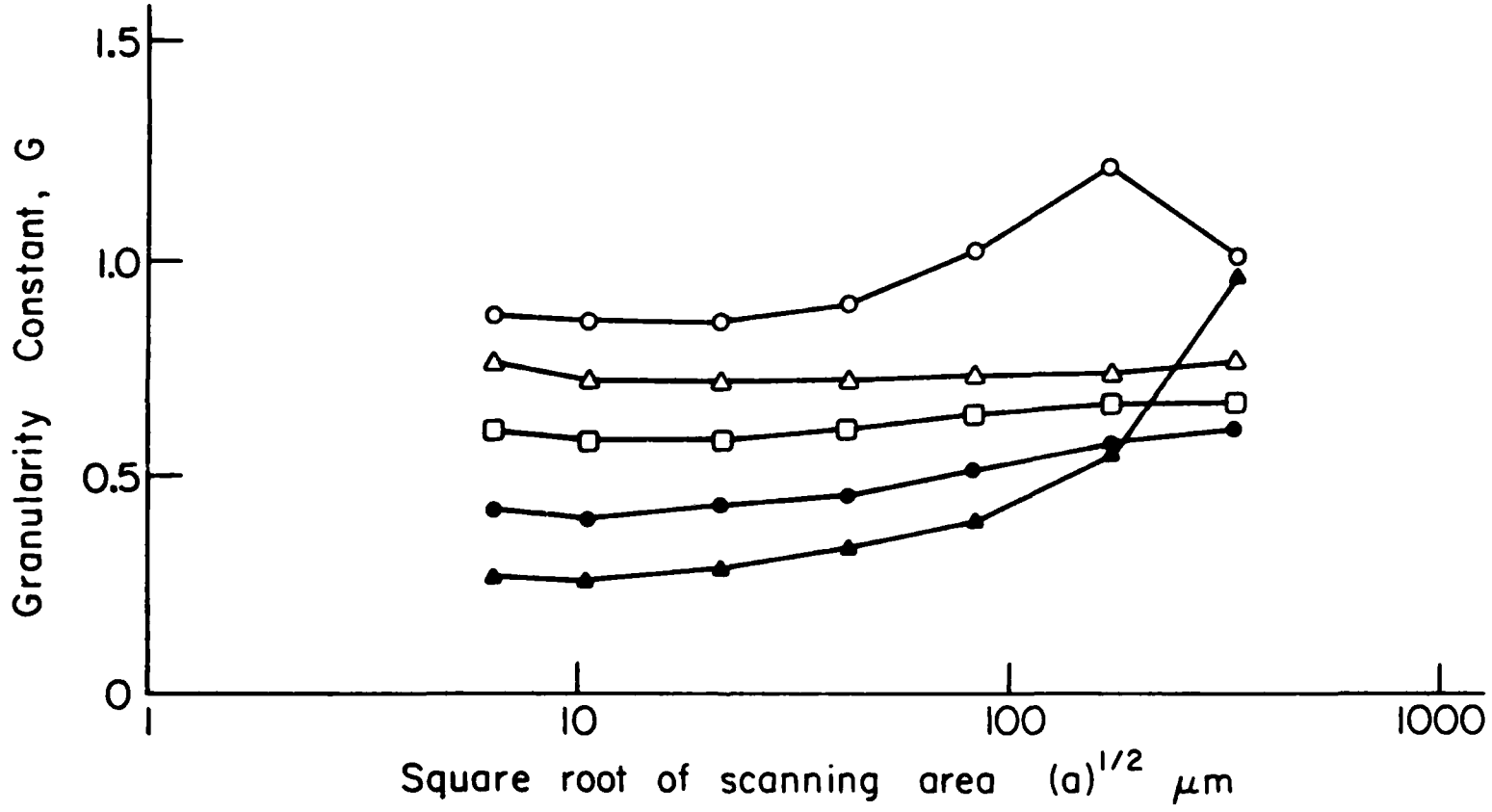


Figure 2. 3-4. Plot of granularity constant G vs. square root of the scanning aperture area for Eastman Fine Grain Release Positive Film. Density levels: ○ 2. 18; △ 1. 44; □ 0. 78; ● 0. 34; ▲ 0. 12.

The data obtained by Higgins and Stultz can also be used to verify the $\frac{1}{2}$ exponent of $\overline{D}$ in eq. (2.1-4), although Higgins and Stultz did not do so in their paper. To do this we rewrite eq. (2.1-4) in the more general form

$$\sigma_D = k(\overline{D})^P \qquad (2.3\text{-}1)$$

where

$$k = \frac{k_2}{k_1} \qquad (2.3\text{-}2)$$

and

$$k_1 = \sqrt{\overline{A}} \qquad (2.3\text{-}3)$$

$$k_2 = \sqrt{.43a} \quad . \qquad (2.3\text{-}4)$$

In the future, throughout this study we will refer to k as the scanning constant. As shown in equations (2.3-3) and (2.3-4), the value of k depends on the constant k_1 by the aperture used for scanning and depends on the grain size constant k_2 by the type of film used. Combining equations (2.2-3) and (2.3-1) we get

$$G = \sqrt{2} \cdot k_2 \cdot (\overline{D})^P \; . \qquad (2.3\text{-}5)$$

Taking the logarithm of both sides of the above equation we have

$$\ln G = P \cdot \ln \overline{D} + C \qquad (2.3\text{-}6)$$

where

$$C = \ln(\sqrt{2}\, k_2) \; . \qquad (2.3\text{-}7)$$

Equation (2.3-6) is known as the Siedentopf [10] law. According to eq. (2.3-6) if several frames of a given film are uniformly exposed and developed, each to a different density, and the granularity constant G for each frame is calculated, then the plot of $\ell n\ G$ vs $\ell n\ \overline{D}$ should follow a straight line with slope P.

To experimentally test these results, we used the Higgins and Stultz data to see if the slope of such a line would be $\frac{1}{2}$ as predicted by eq. (2.1-4). In Figs. (2.3-1) to (2.3-4) at each given density seven values for G are given one for each aperture size. An average of these seven values was obtained and the logarithm of these averages were plotted against the logarithm of the densities. Using least squares techniques, a straight line was fit to the data. The results are shown by solid lines in Figs. 2.3-5 to 2.3-8. A fair degree of deviation from the straight line is observed in Figs. 2.3-7 and 2.3-8, and this is expected considering the deviation from horizontal lines in Figs. 2.3-3 and 2.3-4.

Next a second set of lines were obtained by plotting the logarithm of the densities vs. the logarithm of the average values for G taken not over the entire range of the apertures used in Figs. 2.3-1 to 2.3-4 but only over the three smallest apertures. This was done because the lines in Figs. 2.3-1 to 2.3-4 are nearly horizontal over the smallest three apertures. These new plots are shown by broken lines in Figs. 2.3-5 to 2.3-8. Note that the data now follows a

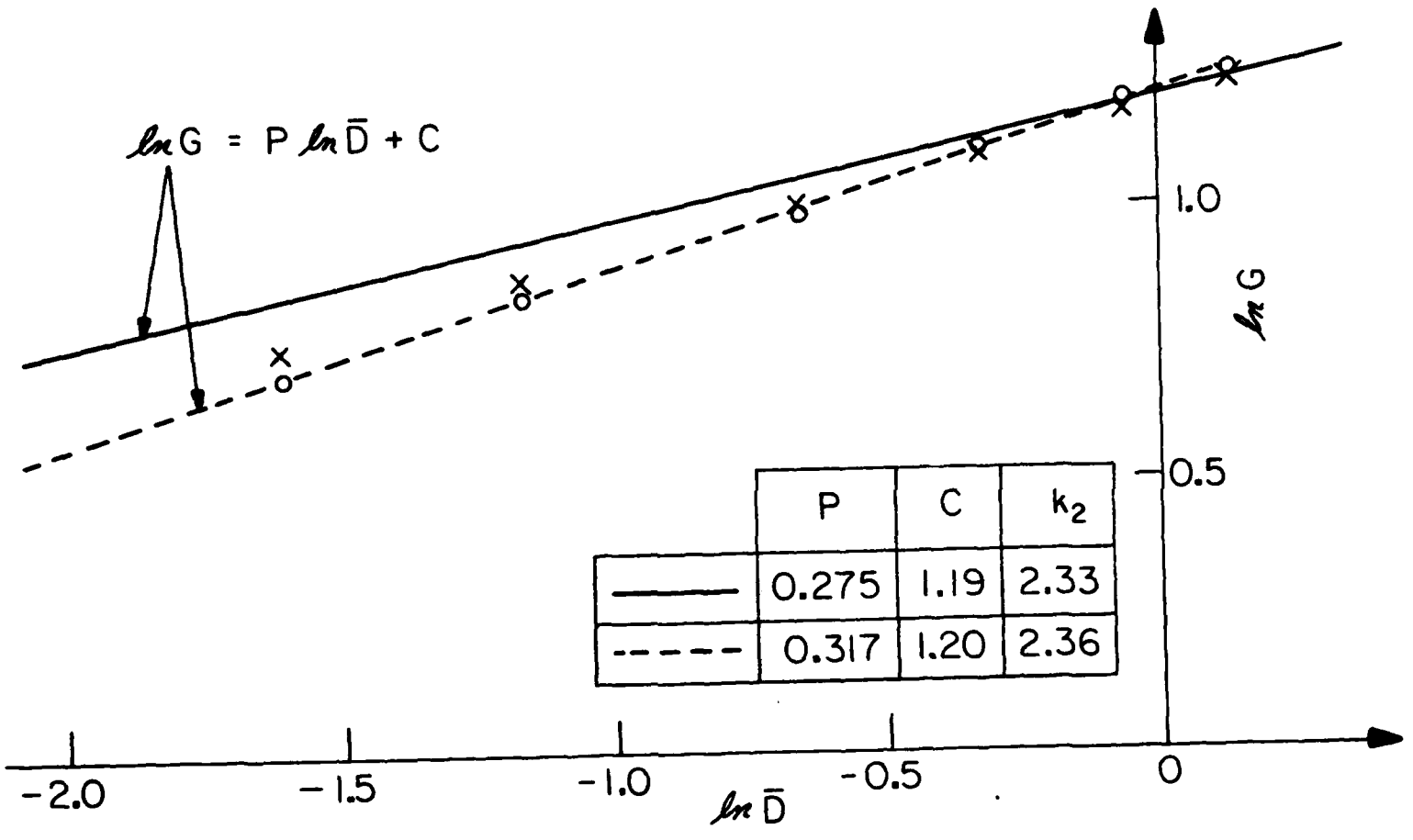
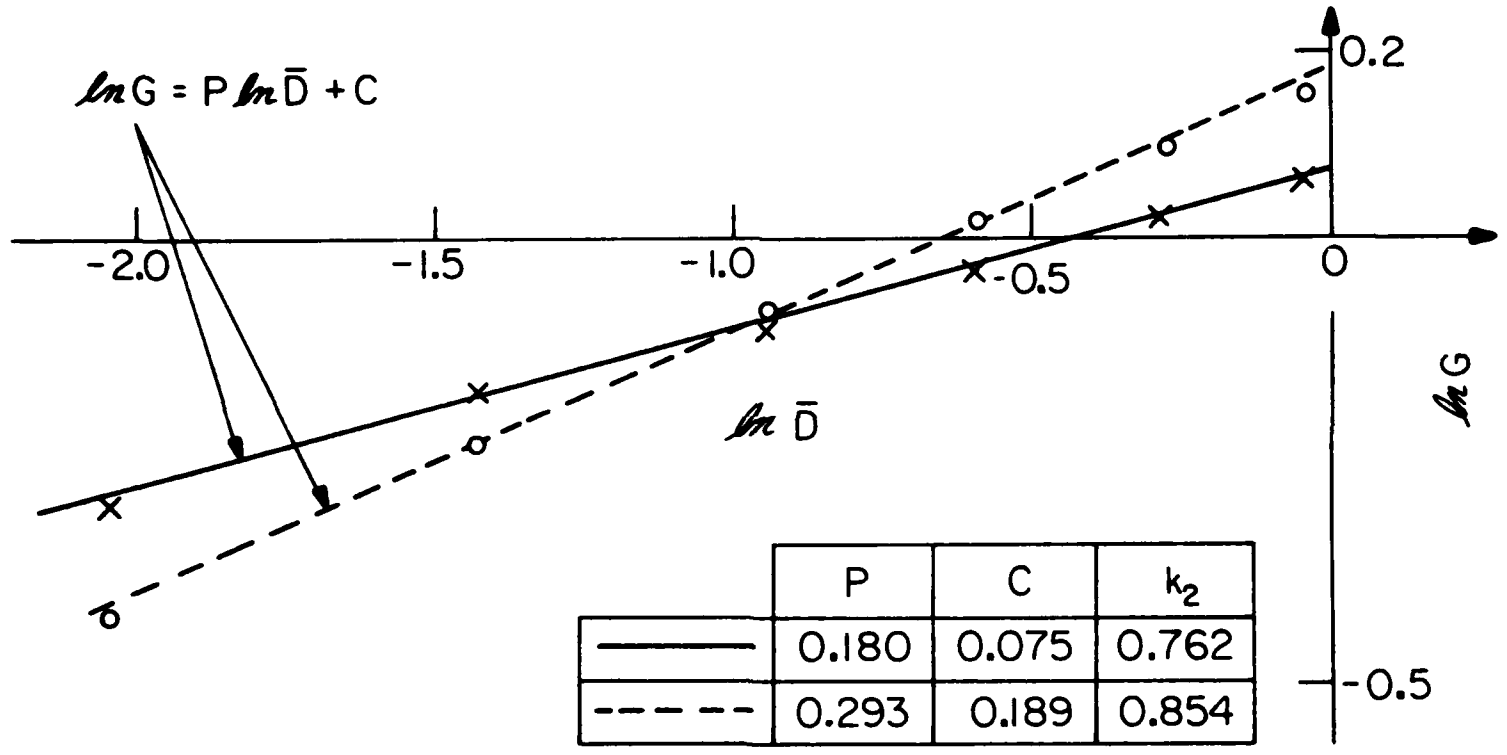


Figure 2. 3-5. Plot of logarithm of the granularity constant G vs. the logarithm of the mean density for Kodak Royal-X Pan film.



	P	C	k_2
———	0.180	0.075	0.762
- - - -	0.293	0.189	0.854

Figure 2. 3-6. Plot of logarithm of the granularity constant G vs. the logarithm of the mean density for Kodak Plus-X negative film.

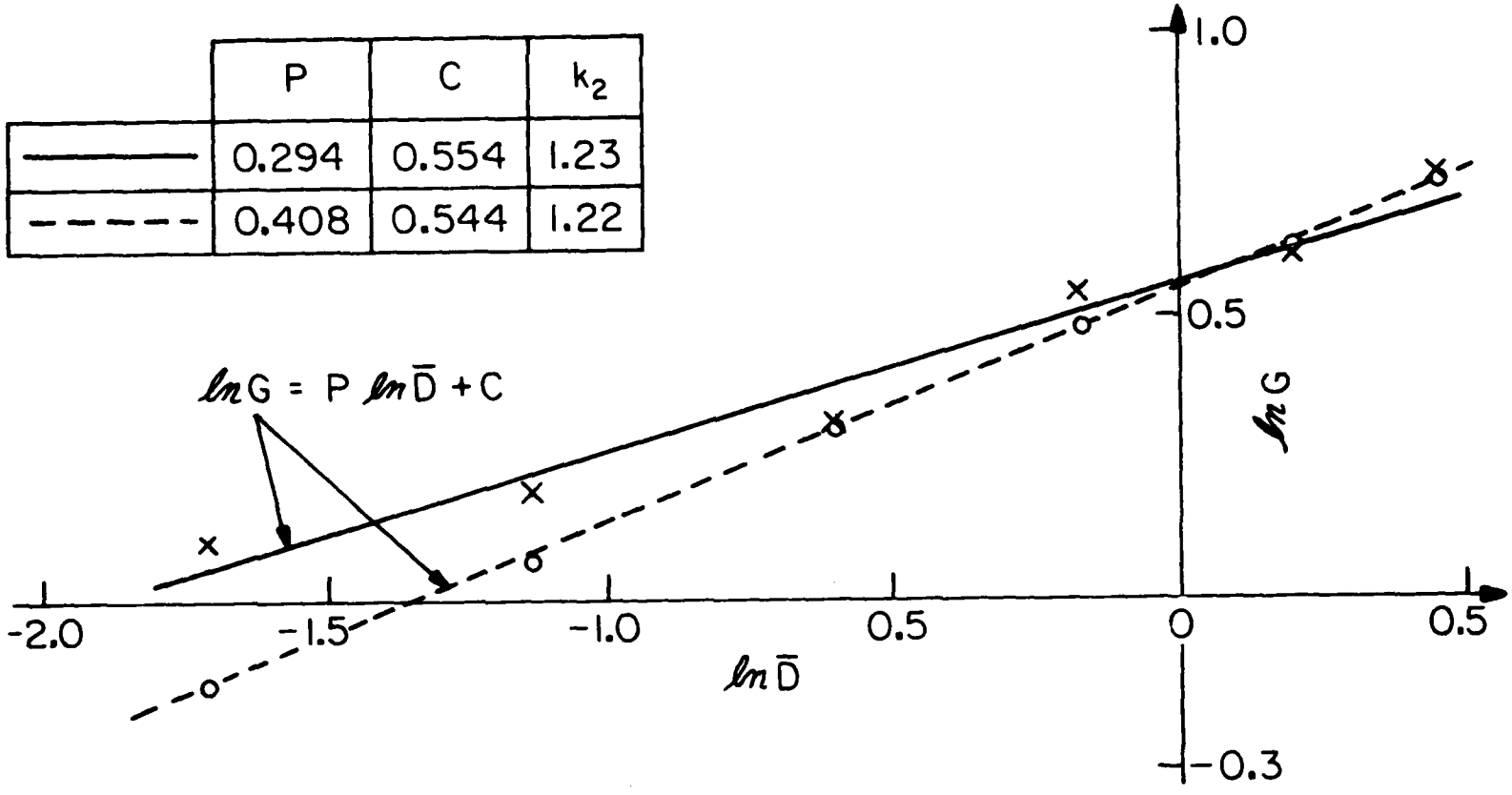
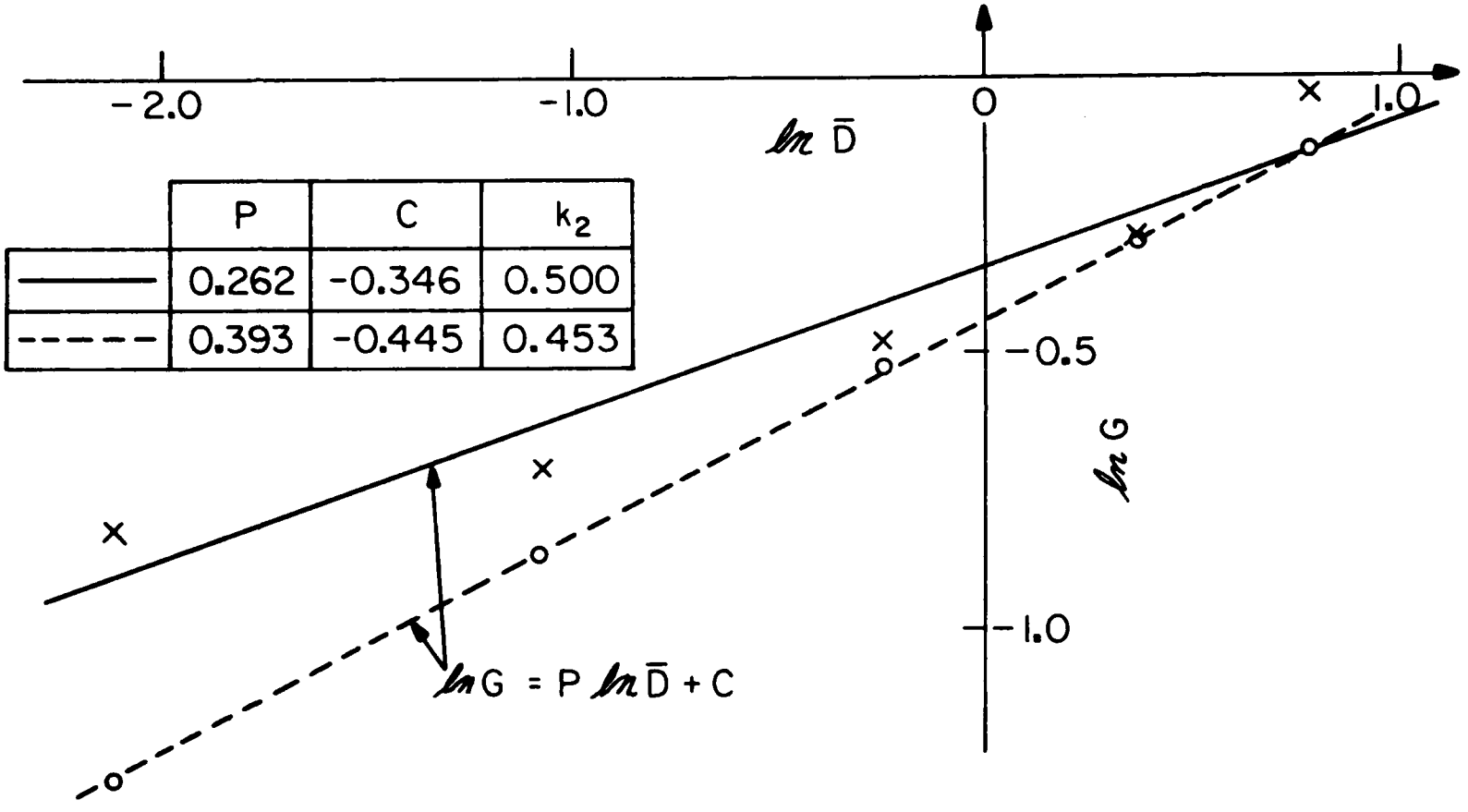


Figure 2. 3-7. Plot of the logarithm of the granularity constant G vs. the logarithm of the mean density for Kodak Panatomic-X Film.



	P	C	k_2
———	0.262	-0.346	0.500
- - - - -	0.393	-0.445	0.453

Figure 2. 3-8. Plot of the logarithm of the granularity constant G vs. the logarithm of the mean density for Eastman Fine Grain Release Positive Film.

straight line very closely in each case. The slopes and intercept points of these lines corresponding to constants P and C, respectively, in eq. (2.3-6) along with the grain size constant k_2 were calculated and are listed in the above figures.

These plots demonstrate that for a reasonable range of apertures, the standard deviation σ_D is of the general form given in eq. (2.3-1). The value of exponent P however appears to be less than 0.5, which was suggested by eq. (2.1-4) through Nutting's model, and is more likely in the range 0.3-0.4. Therefore a more reasonably form for σ_D is

$$\sigma_D = \frac{k_2}{\sqrt{A}} (\overline{D})^{1/3} \qquad (2.3\text{-}8)$$

where k_2 depends on the type of film and is the range 0.45-2.36 for the films used in the Higgins and Stultz study.

Equation (2.1-4) in which exponent of $\overline{D}$ is shown to be $\frac{1}{2}$ was derived based on the assumption that the grains in the developed emulsion all have the same size. Haugh [11] allowed the grain size to follow a distribution and theoretically showed that the exponent P of eq. (2.3-1) is of the form

$$P = \tfrac{1}{2} - \gamma \qquad (2.3\text{-}9)$$

where γ is a positive constant. This then is in agreement with the experimental results just discussed and another justification for taking

$$\sigma_D = k(\bar{D})^{1/3} \qquad (2.3\text{-}10)$$

as the expression for the standard deviation σ_D.

2.4 Distribution of the Optical Density

In previous sections of this chapter we noted that a uniformly exposed and developed film subjectively appears to consist of a uniform shade of gray. However, in examining microscopic regions, one finds that the optical density for these regions fluctuates about the mean density of the film and the standard deviation of the fluctuation has the form given in eq. (2.3-10). In this section we investigate the probability distribution of these fluctuations.

The probability distribution of optical density in microscopic regions of a uniformly exposed and developed film has been studied by photographic scientists for a long time. Most researchers have reported Gaussian or near Gaussian distributions. Some of these reports were made by Bricout [5], Van Kreveld [12] and Selwyn [13]. On the other hand, some other investigators such as Debot [14] and Berwart [15] reported some skewness and asymmetry in their distribution curves. Unfortunately, most of these experiments were done 30 to 40 years ago when accurate instruments were not available. We now discuss the assumption of Gaussianity and show that it should be accepted with some degree of reservation.

In chapter 1, the optical density for a region R in a developed

film was defined as

$$D = -\log_{10} \frac{I_t}{I_i} \qquad (2.4\text{-}1)$$

where I_t was the intensity of the light transmitted through region R when illuminated by a light of constant intensity I_i. Theoretically I_t can be zero, corresponding to infinite optical density. However, as shown in characteristic curve of Fig. 1.2-1, all films reach a saturation point beyond which any increase in exposure will not yield higher density. In practice densities above 2.5 are seldom encountered for ordinary photographic films.

In eq. (2.4-1), I_i is always greater or equal to I_t. Therefore, the optical density D is a nonnegative quantity. One can therefore assume that observed values of optical density are approximately in the range of 0.0-2.5. We now will examine assumptions for Gaussianity and standard deviation as given by eq. (2.3-10) to determine if they are consistent within this density range. We first examine the upper end of this range by assuming a film to be uniformly exposed and developed to a very high density. Equation (2.3-10) suggests that the standard deviation monotonically increases with the mean density. However, when a film is thoroughly exposed, almost all the halide grains will be transformed to silver grains and the film will develop to an almost uniformly opaque material with not much deviation from the mean density. Therefore it appears that

above a certain density eq. (2.3-10) is not valid.

As evident from eq. (2.4-1), measuring optical density by a microdensitometer or other measuring devices, requires measuring I_t. If the film is heavily exposed, I_t will be very small and recording I_t would become exceedingly difficult due to the noise in the measuring device. For this reason it has been difficult to obtain accurate data for a film developed to a very high mean density.

Considering the lower end of the density range and keeping in mind that density is nonnegative, we examine the following example. Suppose a uniformly exposed Kodak Panatomic-X film is scanned witn an aperture of size 10 μm. in diameter. According to Fig. 2.3-7, the grain size constant for this film is approximately 1.2. Thus the standard deviation of the optical density fluctuation when this film is scanned with the above aperture is given by eq. (2.3-8) and is

$$\sigma_D = .12(\overline{D})^{1/3}. \qquad (2.4\text{-}2)$$

The smallest mean density $\overline{D}$ about which density fluctuation can be Gaussian with standard deviation of the form given by eq. (2.4-2) has to be at least $3\sigma_D$ above zero so that negative values for the optical density have a very small probability. From. eq. (2.4-2) we see that this density is approximately $\overline{D} = 0.216$, corresponding to a standard deviation of $\sigma_D = 0.072$. Figure 2.4-1a shows such a distribution. This figure and eq. (2.4-2) point out that the distribution

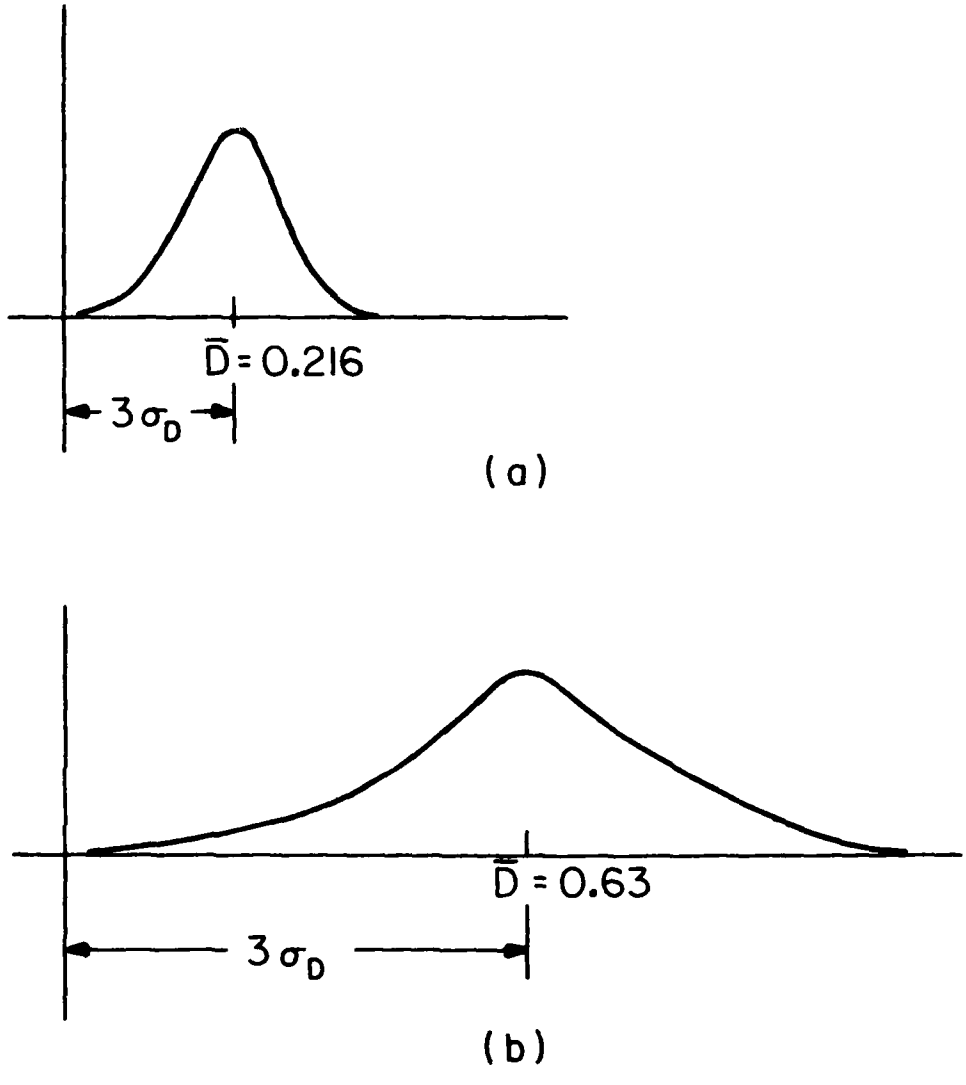


Figure 2.4-1. Distribution of optical density for small values of mean density D.

of the optical density for a uniformly exposed Panatomic-X film scanned with an aperture of size 10 μm. is either not Gaussian or does not have standard deviation of the form given by eq. (2.4-2), when the mean density $\overline{D}$ is in the range [0.0-0.216].

This range will be even wider if a smaller aperture is used for the scanning. If, for example, the diameter of the scanning aperture is only 5 μm., then as shown in Fig. 2.4-1b, the smallest mean density for which $\overline{D}-3\sigma_D$ is nonnegative is 0.63. Using the same argument as above, for this size aperture the density distribution will be non-Gaussian if the mean density is in the range [0.0, 0.63].

If the aperture size is decreased still further then the ratio a/A of the grain size to aperture size will not be small enough for either eq. (2.2-2) or eq. (2.3-8) to hold and the density distribution will not be Gaussian in any range.

Various researchers [14]-[15] have reported that the density distribution at very low levels is skewed. Thus one might model the distribution as a Rayleigh distribution up to some point and by a Gaussian distribution beyond this point. The point would depend on the type of film and the scanning aperture size.

If this choice of modeling appears unattractive because two different distributions are involved, then one might consider modeling by a Rician distribution. The Rician distribution resembles a Rayleigh distribution for a mean close to zero and approaches a

Gaussian distribution as the mean increases [16].

In the early days of granularity studies only 50-100 points were used for calculating the sample standard deviation as well as density distribution due to lack of digital computers and accurate measuring devices. Later on, as many as 1000 points were considered although this still might not have been sufficient to obtain accurate results.

Since those early days, however, the accuracy of measuring devices has been considerably improved. Due to this fact, along with the availability of digital computers, quite a bit can be gained by a more thorough study of the distribution of optical density. This task, however, is not undertaken in this study. In most cases we assume the distribution to be Gaussian.

2.5 A Model for Optical Density in Context of Estimation Theory

As noted before, many of the earlier studies on photographic film were performed about 1900. Most of these studies were done by photographic scientists whose primary goal was analyzing different films and establishing a quantitative measure ranking different emulsions according to their subjective grainy appearance.

Research in the area of film-grain noise, its nature and modeling, reached its height during the late forties to mid-sixties. During this period photographic scientists left a vast volume of publications often resulting in contradictory findings. Unfortunately,

in many of these papers the mathematical notions of linearity and signal dependence are used either too loosely or interchangeably. One goal of this study is to use the theoretical and experimental results on film characteristics obtained by photographic scientists in order to define film-grain noise within the contexts of estimation theory. It is hoped that these models will provide a stepping stone for the eventual restoration of the degraded image.

Earlier in this chapter we noted that film can be viewed as a system in which exposure is the input, optical density is an output and the characteristic function of the film is the transfer function relating the input and the output in macroscopic regions. Thus the optical density in macroscopic regions of a uniformly exposed film can be written as

$$\overline{D}_s = g(E) \qquad (2.5\text{-}1)$$

where E is a spatially uniform input exposure, g is the characteristic function of the film and $\overline{D}_s$ is the optical density registered on the film.

As described in section 1.3, the optical density of a uniformly exposed film examined in very small regions of only a few microns in diameter fluctuates about $\overline{D}_s$, with the fluctuation being near Gaussian with standard deviation of the form given in eq. (2.3-10). Thus we can write

$$\begin{cases} D_o = \overline{D}_s + n \\ n \epsilon N(0, k^2 (\overline{D}_s)^{2/3}) \end{cases} \qquad (2.3\text{-}5)$$

where N(0, 1) is the zero mean, unit variance Gaussian density. In the above equation D_o is the density registered in microscopic regions on film and n is the noise of the photographic film, taken to be a zero-mean Gaussian random variable with standard deviation of $k(\overline{D}_s)^{1/3}$. Equation (2.5-2) implies that D_o is Gaussian distributed with mean $\overline{D}_s$ and standard deviation $k(\overline{D}_s)^{/1/3}$ and statistically matches the measured characteristics of a uniformly exposed film in the first two moments.

While uniformly exposed films are very helpful in studying the statistics of film-grain noise, most images of interest are considerably different. We now consider exposures with spatial variations and see if the model of eq. (2.5-2) also describes the optical density of nonuniformly exposed films.

Let E(x, y) be a deterministic function representing the spatial variation of the exposure on an imaginary plane located just in front of the film to be exposed. Dividing the surface of this plane into N^2 cells of size a x a (see Fig. 2.5-1) and assuming that the spatial variation of exposure is small within each cell, we denote by E_{ij} the exposure level within the cell (i, j).

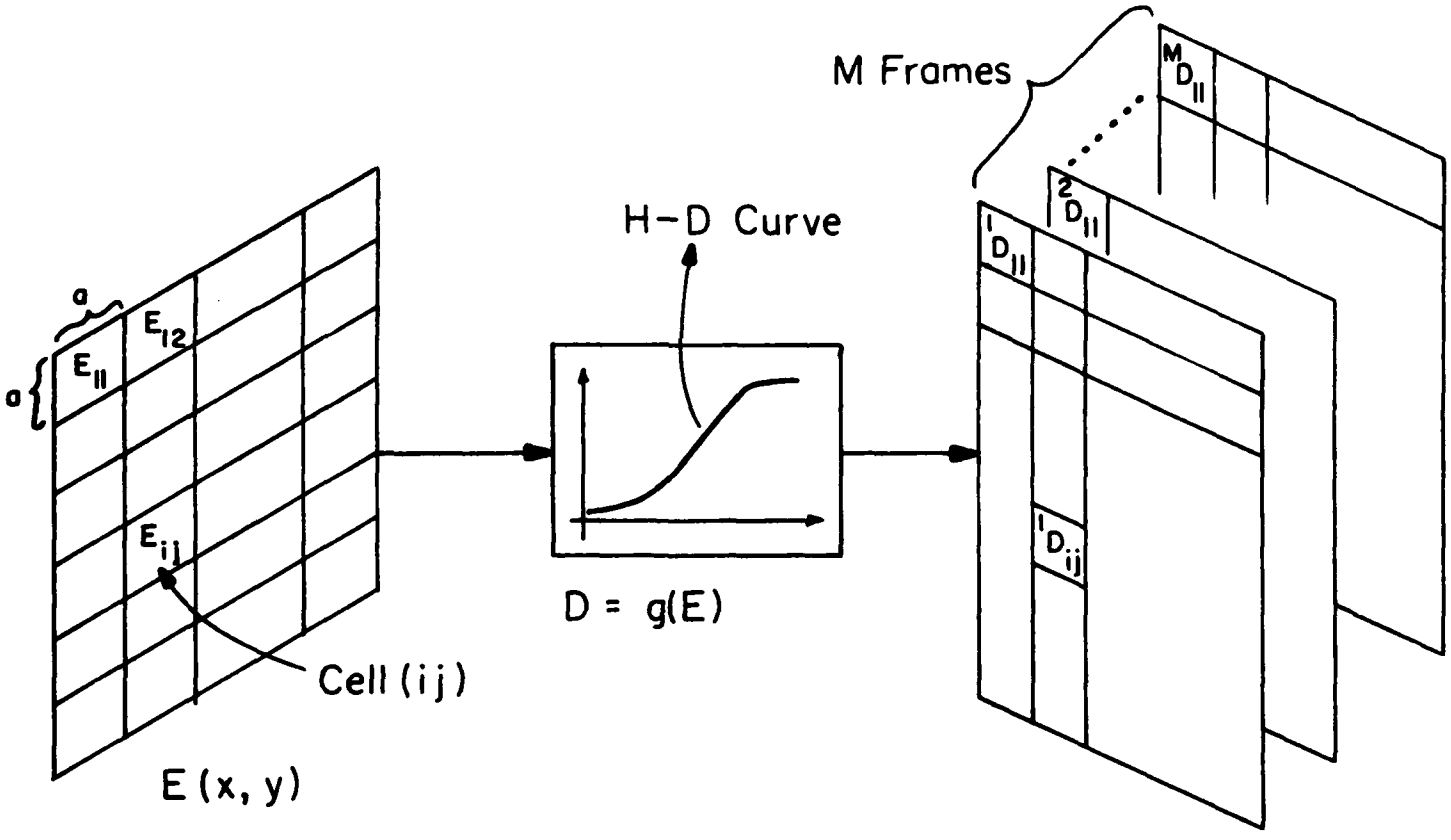


Figure 2.5-1. A model for nonuniformly exposed films.

Suppose M frames of a photographic film of a particular type are exposed with the same exposure $E(x,y)$ and after development their surfaces are divided into N^2 cells of size $a \times a$. We denote by D_{ij}^m the optical density registered in the (i,j) cell of the m^{th} frame. Now, if photographic film were noise free, then D_{ij}^m would have to be identical for $m = 1,\ldots,M$ because the exposure for all the frames is the same. Denoting this common value by $\bar{D}_{ij}$, the input exposure E_{ij} and $\bar{D}_{ij}$ are related through the characteristic curve by

$$\bar{D}_{ij} = g(E_{ij}) . \tag{2.5-3}$$

As we have seen, however, the film noise causes D_{ij}^m to be different for each frame. Since the input exposure was the same for all of the frames, then the M density D_{ij}^m, $m = 1,\ldots,M$ readings can be taken as though they were M readings from a single frame uniformly exposed with exposure E_{ij}.

Letting D_{ij}^m, $m = 1,\ldots,M$ be samples from a random variable D_{ij}, then D_{ij} will be Gaussian distributed with mean $\bar{D}_{ij}$ and standard deviation $k(\bar{D}_{ij})^{1/3}$, where k is a constant depending on the type of the film and the size of the cells. Thus a model reflecting the first two moments of D_{ij} would have the form

$$\begin{cases} D_{ij} = \bar{D}_{ij} + n_{ij} \\ n_{ij} \in N(0, k^2(\bar{D}_{ij})^{2/3}) . \end{cases} \tag{2.5-4}$$

In the above equation $\overline{D}_{ij}$ is nonrandom and n_{ij} represent the randomness (noise) associated with photographic system.

In discussions leading to eq. (2.5-4) we assumed the input exposure $E(x,y)$ to be a deterministic function. The only randomness was due to the noise n introduced by the film. However in most practical cases $E(x,y)$ is also random, so that a general model for the photographic film would have the form

$$D_s = g(E) \tag{2.5-5}$$

$$\begin{cases} D_o = D_s + n \\ n \epsilon N(0, k^2 (D_s)^{2/3}) \end{cases} \tag{2.5-6}$$

where in the subscripts (i,j) are dropped but understood. Here E is the input exposure, D_s is the resulting density in the absence of film-grain noise, n is the noise and D_o is the observed density. The quantities E, D_s, n and D_o are all random variables.

2.6 Linearity vs. Signal Dependence

There is a fair amount of confusion in the literature regarding the so-called multiplicative nature of film-grain noise. From eq. (2.5-6) it is clear that film-grain noise can be modeled as additive noise. However, this noise is signal dependent by virtue of its variance. Alternatively, the additive nature of the noise can be sacrificed to eliminate the signal dependence. In this case optical density can be modeled as

$$\begin{cases} D_o = D_s + k(D_s)^{1/3} \cdot n \\ n \epsilon N(0, 1) \end{cases} \qquad (2.6\text{-}1)$$

where the noise n is Gaussian with zero mean and unit variance and is statistically independent of the signal. Equation (2.6-1) now represents a nonlinear model. Note that the model of eq. (2.5-6) and that of eq. (2.6-1) are identical in the sense that the random variable D_o given D_s has the same first two moments in either model. The difference is that in one model we have additive signal dependent noise whereas in the other model we have nonlinear observation with signal independent noise.

In many estimation literature the term multiplicative noise is strictly used to describe observations of the form

$$y = s \cdot n \qquad (2.6\text{-}2)$$

where y is the observation, s is the signal and n is the noise. For this reason we will not refer to film-grain noise of eq. (2.6-1) as multiplicative noise lest that be misleading.

The model of eq. (2.5-6) might appear easier to work with because of the additivity of the noise. However any advantage gained due to linearity of the model is more than offset by the disadvantage of working with signal-dependent noise. Thus in subsequent chapters we shall use eq. (2.6-1). This model was also suggested by Huang [17] to model the random part of film-grain effects. Figure 2.6-1 shows

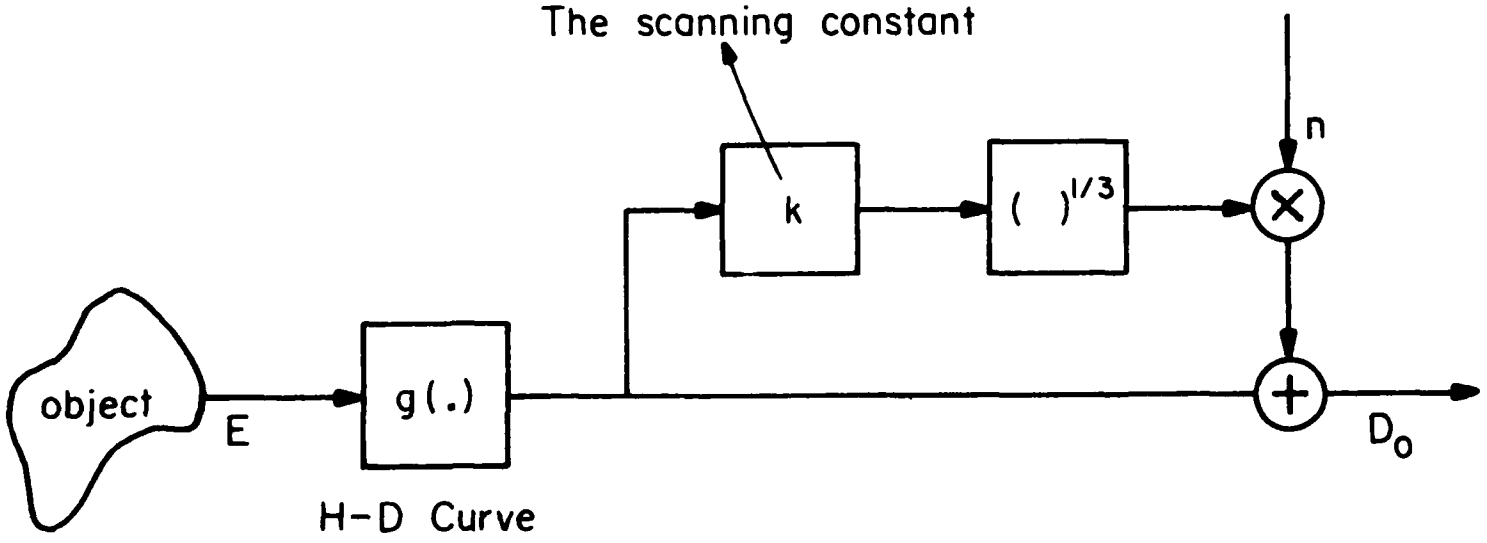


Figure 2.6-1. A model for film-grain noise.

the block diagram for this model.

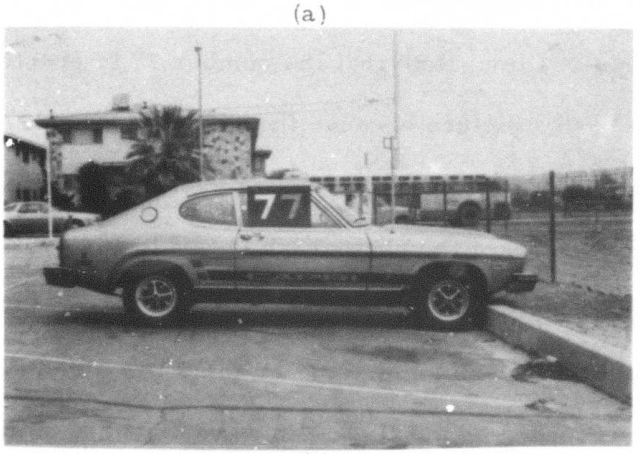
2.7 Evaluation of the Model

The aim of this chapter has been the development of a model for the image forming and recording system which can be used in restoring images degraded by film-grain noise. The reliability of any restoration scheme is directly related to the accuracy by which the underlying mathematical model describes the actual physical process.

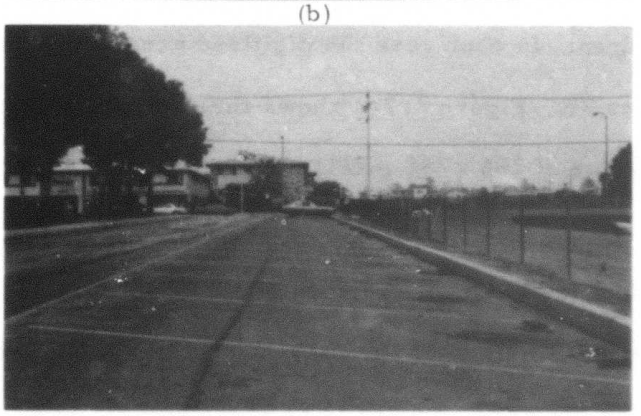
Unfortunately there does not exist any well-defined or tractable fidelity criterion by which various models can be evaluated or compared. For this reason, we decided to test different models according to the following procedure. A target was first photographed and the film developed. Then a section of the film was digitized using a small aperture. Next we simulated the same section of the target in the computer according to the model which was to be evaluated. Since any imaging model must simulate the physical process, we subjectively compared the photographed and simulated images to determine the accuracy of the model. There is no mathematical optimality associated with this type of testing although it is a reasonable procedure, especially for image processing involving the human observer.

Figure 2.7-1a shows the target used for the evaluation. It consists of two figure sevens cut out of white and red cardboard

(a)



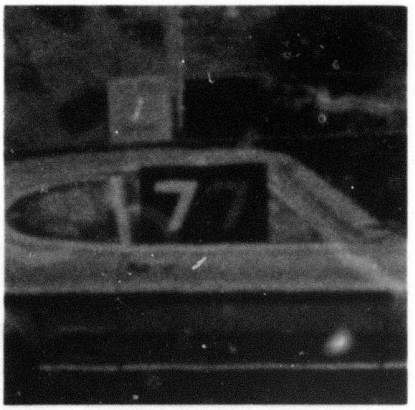
(b)



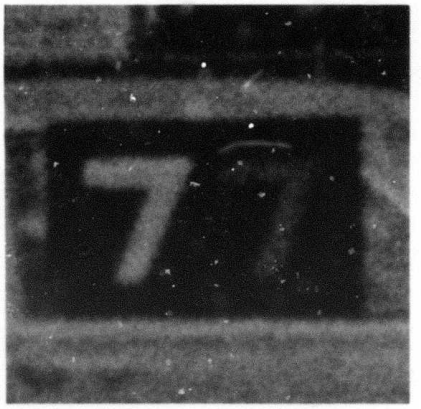
(c)

Figure 2.7-1. The target used for evaluation of the models.

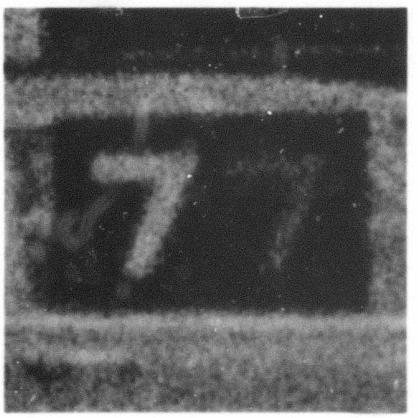
which are then pasted on a black cardboard. Photographing this target on a monochrome film produces an image with three levels. This target was then attached to the side of a car as shown in Fig. 2.7-1b, and photographed at various distances using two different types of film. One was the high speed Kodak Tri-X film and the other was Kodak Panatomic-X, a film with much slower speed. Figure 2.7-1c shows one of the photographs shot at a long distance on a Pan-X film. Note that the number 77 is practically impossible to see in this picture because the target of Fig. 2.7-1a is mapped into a square region of dimensions about 0.5 millimeter on each side on the negative of the photograph shown in Fig. 2.7-1c. This area is much too small for the eye to resolve. Next, the negative of the photograph shown in Fig. 2.7-1c and a negative of exactly the same scene recorded on a Tri-X film were digitized using various small apertures. In each case the digitized area was centered around the number 77. Figure 2.7-2 shows three such digitized images, all consisting of 256 x 256 pixels. Figure 2.7-2a is the Pan-X film digitized with a $5 \times 5\ (\mu m.)^2$ aperture, and Fig. 2.7-2b is the same film digitized with an aperture of size $2 \times 2\ (\mu m.)^2$. As predicted by the denominator of eq. (2.3-8), the image digitized with the smaller aperture appears to be more noisy. Finally, Fig. 2.7-3c shows the Tri-X film digitized with a $2 \times 2\ (\mu m.)^2$ aperture. Once again comparing figures (b) and (c) we see that Fig. (c) is more noisy



(a) Panatomic-X Film digitized with a 5x5 $(\mu m.)^2$ aperture.

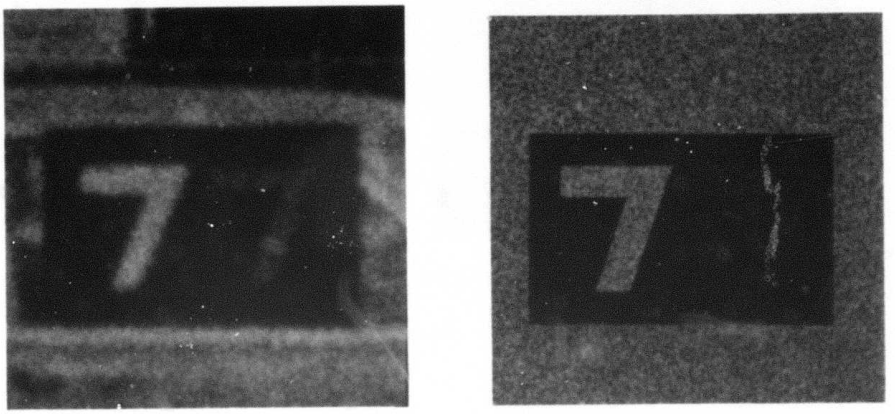


(b) Panatomic-X Film digitized with a 2 x 2 (μm.) aperture.



(c) Tri-X Film digitized with a 2 x 2 $(\mu m.)^2$ aperture.

Figure 2.7-2. Digitized images obtained by scanning two types of film with two different size apertures.



(a) Actual image.

(b) Simulated image

Figure 2.7-3. Comparison of the actual image with the image simulated according to the model of Fig. 2.6-1.

as predicted by eq. (2.3-4). This results because the average grain size a is larger for Tri-X film than for Pan-X film.

We selected Fig. 2.7-2b as one of the test targets to be used in evaluation of the model shown in Fig. 2.6-1. The two images used for the evaluation are shown in Fig. 2.7-3. Figure 2.7-3a is the actual photographed image and is the same image shown in Fig. 2.7-2b. The image simulated in the computer according to the model of Fig. 2.6-1 is shown in Fig. 2.7-3b.

The question now is whether the simulated image gives the same subjective impression as the actual photographed image. The answer is clearly no. The edges of the sevens in the simulated image are very sharp whereas they are quite blurred in the actual image. The noise in Fig. 2.7-3b has the salt and pepper appearance whereas in Fig. 2.7-3a this is not the case. This comparison points out that the model of Fig. 2.6-1 alone is not quite representative of the imaging process. Thus it either needs to be altered or augmented.

2.8 A Model for Image Forming and Recording Process

As discussed in Chapter 1 the process of recording an image on a photographic film is a rather complex process. Not all the degradation is due to film-grain noise. There are other complex optical and chemical effects that contribute to the degradation of the recorded image that were neglected in the model of Fig. 2.6-1. Any accurate modeling should account for these effects as well as the granularity

noise.

Around 1960, Kelly [18] suggested a three stage model which included optical and chemical degradation effects, but ignored the granularity effects. Later Lorber [19] proposed cascading a fourth stage to Kelly's model to account for film-grain noise.

Combining these results with the discussion in the previous seven sections of this chapter, a model of the form shown in Fig. 2.8-1 is suggested. In this figure, $I_1(x, y)$ is the intensity of the light reflected from the object and is incident on the image forming device such as a camera. The first of four linear blurs in this model represents atmospheric degradations such as turbulence as well as limiting effects of the imaging system such as diffraction and aberrations. The point-spread function for this linear blur is $b_1(x, y)$. $I_2(x, y)$, the output of this initial block is the intensity distribution of the light that actually strikes the film in the camera.

The second linear blur models optical diffusion effects such as scattering and halation, during the formation of latent image. This blur has a point-spread function $b_2(x, y)$. The middle segment of this model is exactly the same as the model of Fig. 2.6-1, and represents the nonlinear characteristic function g as well as the nonlinear signal-independent film-grain noise.

The third linear blur with point spread function $b_3(x, y)$ accounts for degradations such as adjacency effects [19] which arise during

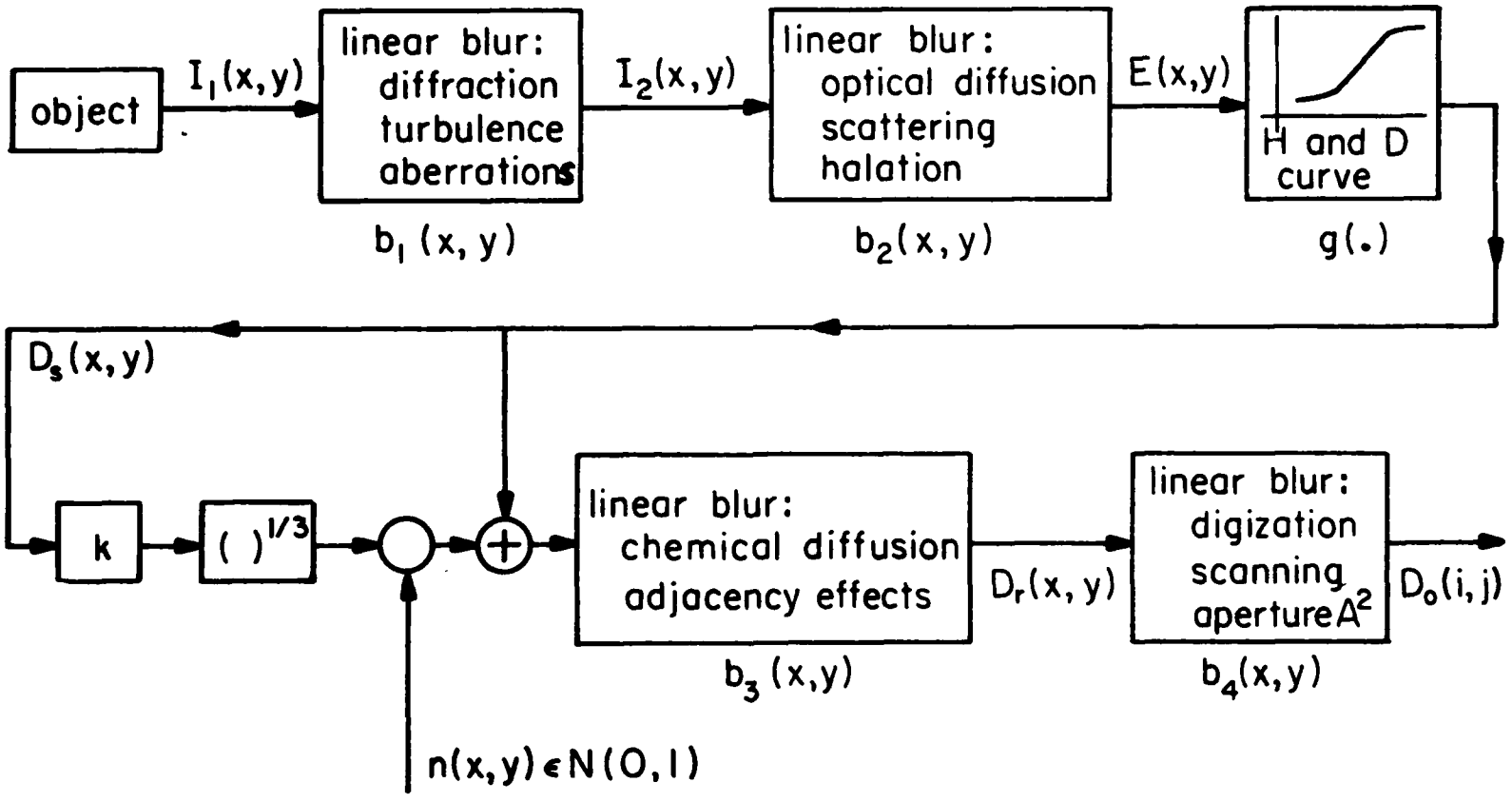


Figure 2.8-1. A model for the imaging system.

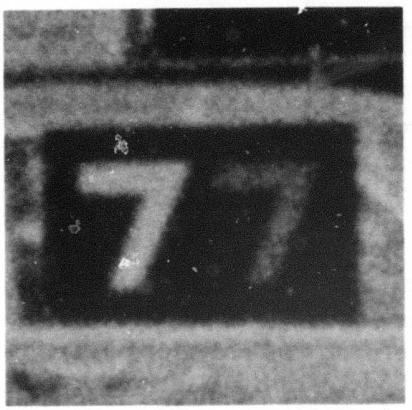
the development of the film. Finally, the digitization process is modeled by the fourth linear blur. The aperture of the scanner is represented by point-spread function $b_4(x,y)$. The output of this model is the observed digitized image $D_o(i,j)$.

To evaluate this model once again we simulated the image recording process according to it and compared the simulated image to an actually photographed image. The results are shown in Fig. 2.8-2.

Figure 2.8-2a is the actual image and Fig. 2.8-2b is the simulated image according to model of Fig. 2.6-1 accounting only for granularity noise. Figure 2.8-2c is the image simulated according to the suggested model of Fig. 2.8-1.

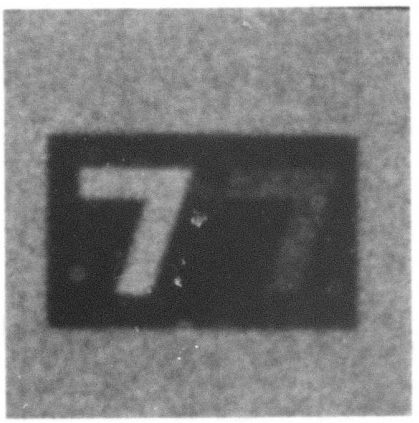
Comparing figures (a) and (c) it is clear that the two images are very similar. Thus it is our conclusion that the model suggested in Fig. 2.8-1 is a very reasonable model for describing the imaging system.

Figure 2.8-3 describes the detailed procedure for simulating 2.8-2c. We first started with a simulated ideal target which resembled Fig. 2.7-1a in size, shape and density. Next, this 256 x 256 image was blurred using a Gaussian blur of window 11 x 11 pixels and standard deviation 2.0. Nonlinear Gaussian noise was added to the blurred target in accordance with the model of Fig. 2.8-1 with the scanning constant chosen to be 0.3. Finally, the noisy



(a) The actual image

(b) Simulated image accounting for film-grain noise only



(c) Simulated image accounting for all degradations

Figure 2.8-2. Comparison of the actual image with images simulated according to the models of Fig. 2.6-1 and 2.8-1.

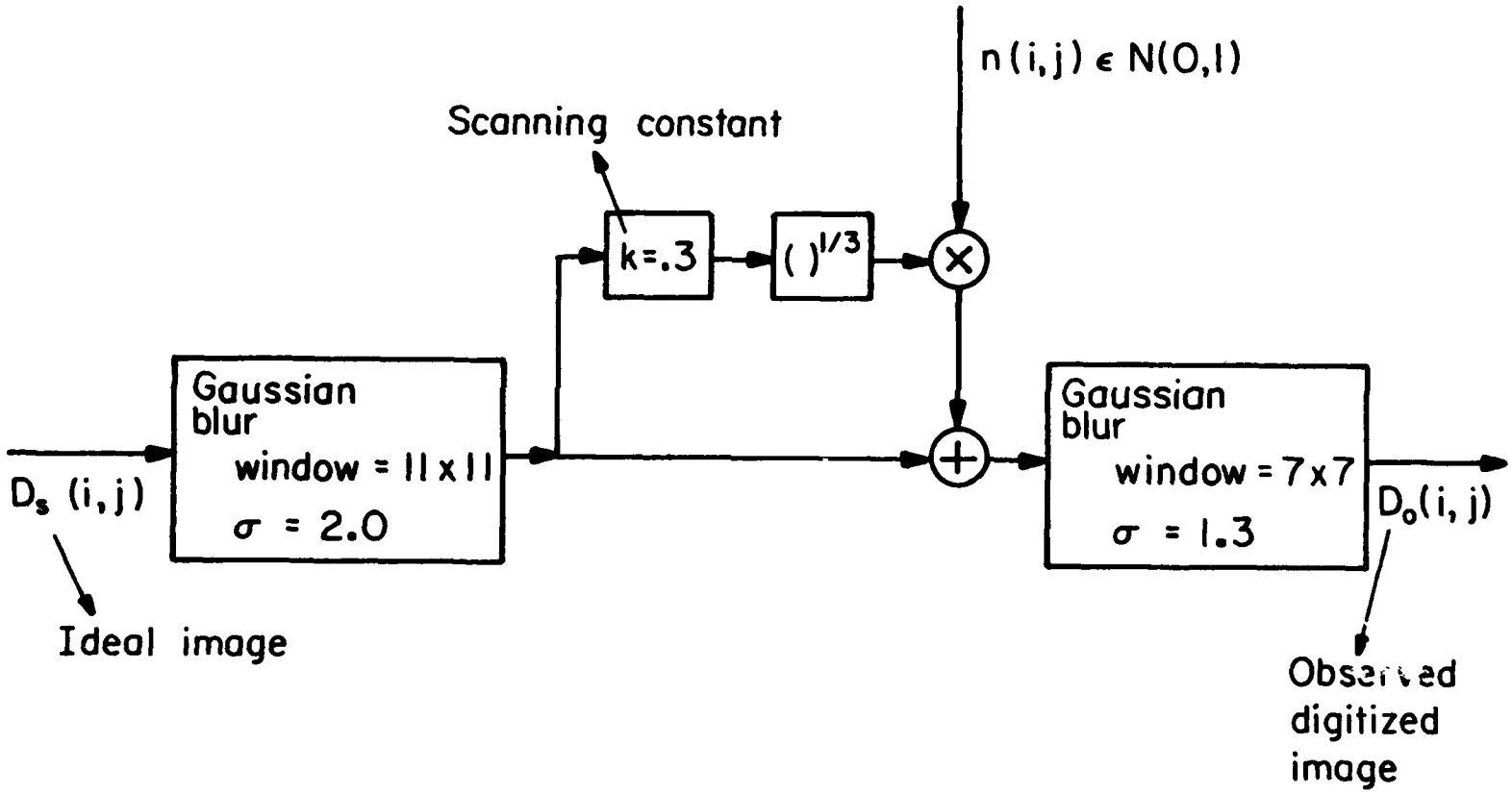


Figure 2.8-3. Procedure describing simulation of Fig. 2.8-2c.

image was once again blurred with a Gaussian blur of window size 7 x 7 pixel and standard deviation 1.3 to obtain the final results.

It should be noted that if we wanted to simulate an image which was digitized with a larger aperture, we would have reduced the window size of the last blur function in order to get a better resemblance with the actual image. This is done because adjacency effects are not as noticeable in images digitized with large apertures.

REFERENCES

[1] P. G. Nutting, "On the Absorption of Light in Heterogeneous Media," Phil. Mag., Vol. 26, p. 423, 1913.

[2] J. C. Dainty and R. Shaw, Image Science, Academic Press, London, p. 41, 1974.

[3] C. E. K. Mees and T. H. James, The Theory of the Photographic Process, 3rd Ed., Macmillan, N.Y., 1966.

[4] W. Thomas, SPSE Handbook of Photographic Science and Engineering, John Wiley & Sons, p. 939, N.Y., 1973.

[5] P. Bricout, Compt. Rend., Vol. 197, p. 1202, 1933.

[6] E. W. H. Selwyn, "A Theory of Graininess," The Phot. J., Vol. 75, p. 571, 1935.

[7] R. Debot, Research (London), vol. 2, p. 139, 1949.

[8] G. Hansen and P. H. Keck, Z. Wiss. Photo., vol. 37, pp. 86-99, 1938.

[9] G. C. Higgins and K. F. Stultz, "Experimental Study of rms Granularity as a Function of Scanning-Spot Size," J. Opt. Soc. Am., vol. 49, p. 925, 1959.

[10] H. Siedentopf, "Concerning Granularity, Resolution and the Enlargement of Photographic Negatives," Physik Zeit., vol. 38, p. 454, 1937.

[11] E. F. Haugh, "A Structural Theory for Selwyn Granularity Coefficient," The J. Phot. Sci., Vol. 11, p. 65, 1963.

[12] A. Van Kreveld, J. Opt. Soc. Am., Vol. 26, p. 170, 1936.

[13] E. W. H. Selwyn, "Experimental on Nature of Graininess," The Phot. J., p. 513, Aug. 1939.

[14] R. Debot, Research (London), Vol. 2, p. 341, 1949.

[15] L. Berwart, Bull. Soc. Roy. Sci. Liege, Vol. 32, p. 794, 1963.

[16] J. Omura and T. Kailath, "Some Useful Probability Distribution," Technical Report No. 7050-60, Stanford Univ., Sept. 1965.

[17] T. S. Huang, "Some Notes on Film-Grain Noise," Appendix 14, "Restoration of Atmospherically-Degraded Images," NSF Summer Study Rept., Woods Hole, Mass., pp. 105-109, 1966.

[18] D. H. Kelly, "Systems Analysis of the Photographic Process I. A Three-Stage Model," J. Opt. Soc. Am., Vol. 50, p. 269, 1960.

[19] H. W. Lorber, "A Theory of Granularity and Bleaching for Holographic Information Recording," IBM Res. Dev., Vol. 14, No. 5, Sept. 1970.

Chapter 3

ESTIMATION

In this chapter the restoration of images degraded by film-grain noise is considered within the context of estimation theory. In section 3.1 we investigate the applicability of some existing estimators to film-grain noise restoration. Next, in section 3.2, a linear minimum mean square error (LMMSE) filter based on non-stationary first order statistics of an image is developed. Finally in section 3.3, this filter is applied to one-dimensional and two-dimensional signals and its performance is compared with conventional discrete Wiener filter.

3.1 Survey of Some Estimators

To compare and study the relevance of some existing estimators to the problem of film-grain noise restoration, let us establish some notation. Let us denote the signal to be estimated by the random sequence $x_1, x_2, \ldots, x_k, \ldots, x_N$ where x_k is the signal at time or position k. Similarly let $y_1, y_2, \ldots, y_k, \ldots, y_N$ denote the observation sequence with y_k the observation at time or position k. Usually y_k is a function of x_k and a noise term n_k which has caused the degradation.

We can generally write

$$y_k = f(x_k, n_k, k) \qquad (3.1\text{-}1)$$

where f is a deterministic function. Equation (3.1-1) is often referred to as the observation model and as we shall discuss, the implementation of the estimator designed to estimate x_k based on observation y_k can vary greatly in complexity depending on the nature of the function f.

One can classify various estimators as either batch processors or recrusive estimators. A batch processor processes all of the observations $y_1, \ldots, y_N$ at the same time and outputs the estimates of $x_1, \ldots, x_N$ all at once. On the other hand, a recursive estimator provides an estimate for x_k based only on a small number of observations and past estimates. The procedure is then repeated to produce an estimate for x_{k+1} and so on.

If can be shown [1] that for a large class of cost functions the best estimate (the best in terms of minimizing the average cost) of x_k using observations $y_1, \ldots, y_\ell$ is given by

$$\hat{x}_k = E[x_k | y_1, \ldots, y_\ell] \qquad (3.1\text{-}2)$$

where E denotes the ensemble average operator. In the above equation if $\ell = k$ then $\hat{x}_k$ is called the filtered estimate. If $\ell < k$ the estimate is called the predicted estimate and if $\ell > k$, $\hat{x}_k$ is called the interpolated estimate of x_k. In this last case where the estimator

uses observations in the future to estimate the signal at the present, the estimator is called non-causal. Obviously if a high degree of correlation exists within the signal samples and/or within the noise samples then a non-causal estimator always gives a better estimate than a causal estimator because it uses more of the available information. Clearly a batch estimator as we have defined it in this section is a non-causal estimator. In most situations in image processing, the quantity of data to be processed is enormous and there are difficulties in implementing a batch estimator. To illustrate this point, consider a digitized image consisting of 256 by 256 pixels. A general batch estimator will have to process over 65,000 observations at the same time to produce an estimated image. With the size of most digital computers this causes storage and computational difficulties. By comparison, a recursive estimator may or may not be causal, but in either case is usually easier to implement because it uses a relatively small amount of data at any given time to produce an estimate.

Both recursive and batch estimators can be subdivided into linear and nonlinear estimators. The linear estimator produces an estimate which is a linear combination of the observations.

As stated earlier, the optimal Bayesian estimate for a large class of cost functions is given by

$$\hat{\bar{x}} = E[\bar{x}|\bar{y}] \qquad (3.1\text{-}3)$$

where $\bar{y}$ denotes the observation to be used for the estimate and $\bar{x}$ is either the signal at time k or the entire signal sequence $x_1, \ldots, x_N$ depending on whether the estimator is recursive or a batch estimator. Using the definition of ensemble average we can rewrite equation (3.1-3) as

$$\hat{\bar{x}} = \int_{\bar{x}} \bar{x} P(\bar{x}|\bar{y}) d\bar{x} \qquad (3.1\text{-}4)$$

where $P(\bar{x}|\bar{y})$ is the a posteriori probability density of $\bar{x}$.

One can show that the estimate given by eq. (3.1-3) is the best estimate regardless of the functional relationship between the observation, the signal and the noise. However, if the function f in the observation model of eq. (3.1-1) is nonlinear in signal and noise then the calculation of the probability density function required to carry out the estimate of eq. (3.1-4) can become very difficult. The general problem of implementing the optimal estimator using nonlinear observations still remains unsolved. The most recent work done in this area is by Naraghi [2] who proposes a sub-optimal algorithmic estimation method applicable to nonlinear observations.

Even with many simplifying assumptions, estimation of images degraded by film-grain noise, still involves nonlinear observations. As discussed in the last chapter, D_o, the observed optical density of the degraded image is given by

$$D_o = D_s + kD_s^P \cdot n \qquad (3.1\text{-}5)$$

where D_s denotes the recorded image in the absence of noise. In the above equation we have ignored all the optical and chemical degradations as well as the effect of the nonlinear H and D curve. We can see that even for this very simplified case the observation is quite nonlinear and the optimal estimate of D_s as given by $E[D_s|D_o]$ would be difficult to implement. For this reason various investigators have tried to find alternative solutions to film-grain noise restoration. In the remainder of this section we consider some of the more recent approaches to this problem to provide a background.

1. Maximum A Posteriori (MAP) Estimator

Recently Hunt [3] has considered image restoration based on a model given in Fig. 3.1-1. This model is a much simplified version of the model shown in Fig. 2.8-1 of Chapter 2. Here the signal dependence of the noise is ignored and the main emphasis is placed on the nonlinear nature of the function $g(\cdot)$. From Fig. 3.1-1 the vector observation model

$$\bar{y} = g([H]\cdot\bar{x}) + \bar{n} \qquad (3.1\text{-}6)$$

can be written. By ignoring the signal dependence of the noise, we see that the nonlinearity in the observation model of eq. (3.1-6) does not include any cross-multiplication term involving the signal and the noise, as was the case in eq. (3.1-5), but rather is completely in

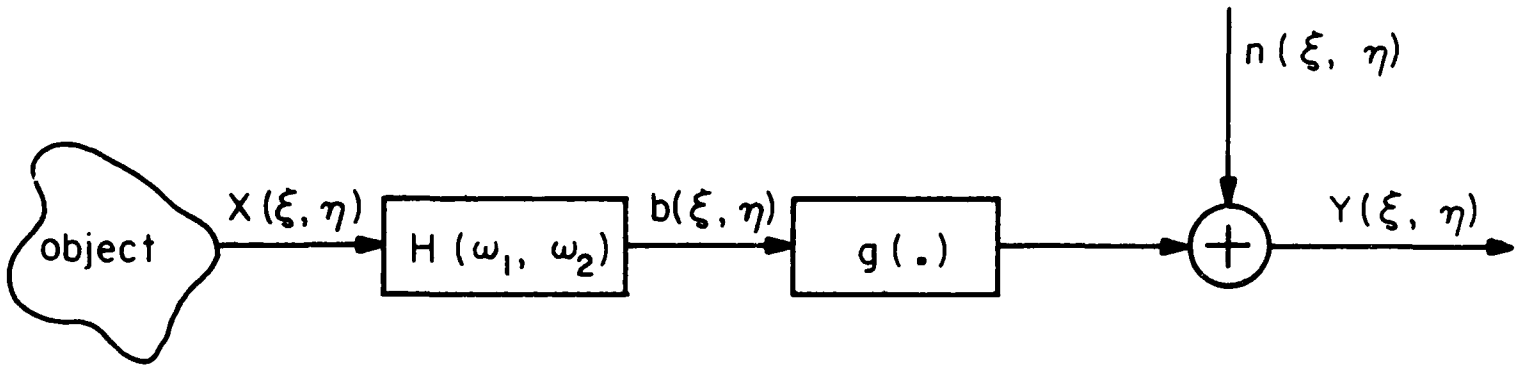


Figure 3.1-1. The imaging model used in the derivation of the MAP estimator.

terms of the signal component only.

Hunt argues that the Bayesian MMSE estimate of $\bar{x}$ given by $E[\bar{x}|\bar{y}]$ involves calculating the a posteriori density $P(\bar{x}|\bar{y})$, which is not simple due to nonlinear function $g(\cdot)$, and instead decides on the MAP estimate of $\bar{x}$. This estimate is the mode of the a posteriori density. Unlike $E[\bar{x}|\bar{y}]$, which is the mean of $P(\bar{x}|\bar{y})$, the MAP estimate does not explicitly require knowledge of $P(\bar{x}|\bar{y})$.

In deriving his estimator equations, Hunt assumes that the noise vector $\bar{n}$ in (3.1-6) is wide-sense stationary and multivariate Gaussian with stationary variation about a nonstationary mean $\bar{x}_m$. Thus, the first two moments of $\bar{x}$ can be expressed as:

$$E\bar{x} = \bar{x}_m \tag{3.1-7}$$

$$E[\bar{x}-\bar{x}_m][\bar{x}-\bar{x}_m]^T = R_{\underline{\underline{xx}}} \tag{3.1-8}$$

where $\bar{x}_m$ is a variable vector and $R_{\underline{\underline{xx}}}$ is a Toeplitz matrix. This is a very reasonable description of most images and is superior to the assumption of wide sense stationary [4].

The MAP estimate as derived by Hunt has the form

$$\hat{\bar{x}}_{MAP} = \bar{x}_m + R_{\underline{\underline{xx}}} H^T S_b R_{nn}^{-1} [\bar{y} - g(H \cdot \hat{\bar{x}}_{MAP})] \tag{3.1-9}$$

where

$$S_b = \begin{bmatrix} \left.\frac{\partial g(u)}{\partial u}\right|_{u=\hat{b}_1} & & & 0 \\ & \left.\frac{\partial g(u)}{\partial u}\right|_{u=\hat{b}_2} & & \\ & & \ddots & \\ 0 & & & \left.\frac{\partial g(u)}{\partial u}\right|_{u=\hat{b}_N} \end{bmatrix} \qquad (3.1\text{-}10)$$

and where

$$\hat{\bar{b}} = [H]\hat{\bar{x}}_{MAP} \qquad (3.1\text{-}11)$$

Equations (3.1-9)-(3.1-11) are a set of nonlinear vector equations requiring a great deal of effort to solve. For a 100 x 100 image, $R_{\underline{\underline{xx}}}$ is $10^4 \times 10^4$ matrix. Hunt uses Picard's iteration method to iteratively solve eq. (3.1-9).

An obvious drawback to Hunt's estimate is the over simplification of his model as given in Fig. 3.1-1 on which the estimate is based. In situations when the operation is on a linear region of the H and D curve (see Fig. 1.2-1 of Chapter 1) the nonlinear function g of eq. (3.1-6) can be well approximated by a straight line. Hunt's emphasis in the observation model is on the nonlinearity of this function rather than on the signal dependence of the noise. To obtain a solution in a reasonable time, the MAP estimator requires a very large, fast digital computer.

2. Wiener Filter

The Wiener filter in the Fourier domain is a classical linear estimation technique applicable to stationary signals. Walkup and Choens [5] use the principle of Wiener filter along with the observation model

$$D_o(\xi, \eta) = D_s(\xi, \eta) + kD_s^{1/3}(\xi, \eta) \cdot n(\xi, \eta) \qquad (3.1\text{-}12)$$

To derive a linear minimum-mean-square error filter given by

$$W(\omega_1, \omega_2) = \frac{\Phi_{D_s D_s}(\omega_1, \omega_2)}{\Phi_{D_s D_s}(\omega_1, \omega_2) + k^2 E[D_s(\xi, \eta)^{2/3}]} \qquad (3.1\text{-}13)$$

where $\Phi_{D_s D_s}$ is the spatial power spectrum of the signal, E is the expectation operator and n is taken to be white noise with zero mean and unit variance.

The filter W given by the above equation assumes images to have stationary first and second order statistics. This assumption is not very reasonable, particularly for first order statistics. In addition, no optical and chemical degradations are included in the observation model of eq. (3.1-12) and filter W will only filter out the noise and will not compensate for other degradations.

3.2 Optimal Discrete LMMSE Filter

The object of this section is to develop an optimal discrete linear minimum-mean-square error filter. Before doing so however we would like to list some desirable features for such a filter:

a) As discussed in chapter 2 and again in section 3.1, the observed optical density can be modeled by

$$D_o = D_s + kD_s^P \cdot n \qquad (3.2\text{-}1)$$

if only the effects of film-grain noise are considered and other degradation factors are ignored. Note that even though the noise n in the above model is statistically independent of the signal D_s, the deviation of the observation D_o from the signal D_s is very much dependent on the signal level through the term $kD_s^P \cdot n$. This deviation is more pronounced when the signal level is high and is less severe when the signal level is low. Thus the filter designed to estimate D_s based on the observation D_o should be aware of this fact. In other words the filter should be adaptive in the sense that it should change its characteristics to account for the presence of higher noise when it is operating in a region where the expected signal level is high.

b) As pointed out in the last chapter, the process of recording an image on film involves optical and chemical degradations in addition to film-grain noise. We would therefore like the restoration filter to compensate for these degradations while filtering out the noise.

c) Most existing restoration filters are designed based on the assumption that images are characterized by stationary first and second moments. As discussed before, this is not a very realistic

assumption, particularly in the first moment. To demonstrate this, Cannon and Hunt give the following argument [4]: Suppose one uses several thousand pictures from Californis drivers' licenses in order to calculate an ensemble mean image. Clearly such an ensemble mean image would not consist of a uniform shade of gray indicating stationarity. More likely, the mean image would consist of an elliptical region where the face is expected to be and probably some dark spots where the eyes, nose and other facial features are expected to be. Thus images are generally non-stationary in first order statistics, and we would like to utilize this in the derivation of the filter.

d) Since we are to process digitized images on digital computers we would like the filter to be discrete.

The first two of the above desired properties can be immediately incorporated in the restoration filter. In the model of Fig. 2.8-1 for the imaging system, the first and the last segments of the model accounts for various degradations other than noise. In the middle segment the nonlinear noise is inserted to reflect the signal-dependence term in eq. (3.2-1). Thus if we derive the restoration filter in accordance with this model, then both the a and the b features mentioned in this section will be incorporated in the filter automatically.

To obtain a discrete filter, we must sample the continuous model

of Fig. 2.8-1. Figure 3.2-1 shows the block diagram of the equivalent discrete model. In this figure $\bar{I}$ is a column vector of size N representing the incident intensity which is the input to the imaging system. The M x N blur matrix H_1 represents the combined effects of linear space-invariant blurs $b_1(x, y)$ and $b_2(x, y)$ of the continuous model. It is assumed that the operation is on the linear region of H and D curve and thus the nonlinear function $g(\cdot)$ of Fig. 2.8-1 is approximated by a straight line. In the discrete model the slope of this line is taken to be α and constant vector $\bar{B}$ contains the intercept point of this line for its elements. Vector $\bar{N}$ represents the noise. It is multiplied by a M x M matrix F which is diagonal and has the form

$$[F] \triangleq k \begin{bmatrix} (d_e^1)^P & & & 0 \\ & (d_e^2)^P & & \\ & & \ddots & \\ 0 & & & (d_e^M)^P \end{bmatrix} \qquad (3.2\text{-}2)$$

where k is the scanning constant and d_e^j is the j^{th} element of vector $\bar{D}_e$. Noise $\bar{N}$ is taken to be white, multivariate Gaussian whose first two moments are given by

$$E(\bar{N}) = 0 \qquad (3.2\text{-}3)$$

$$C_{NN} = E(\bar{N} \cdot \bar{N}^T) = [I] \qquad (3.2\text{-}4)$$

where $[I]$ denotes the identity matrix and E is the ensemble average

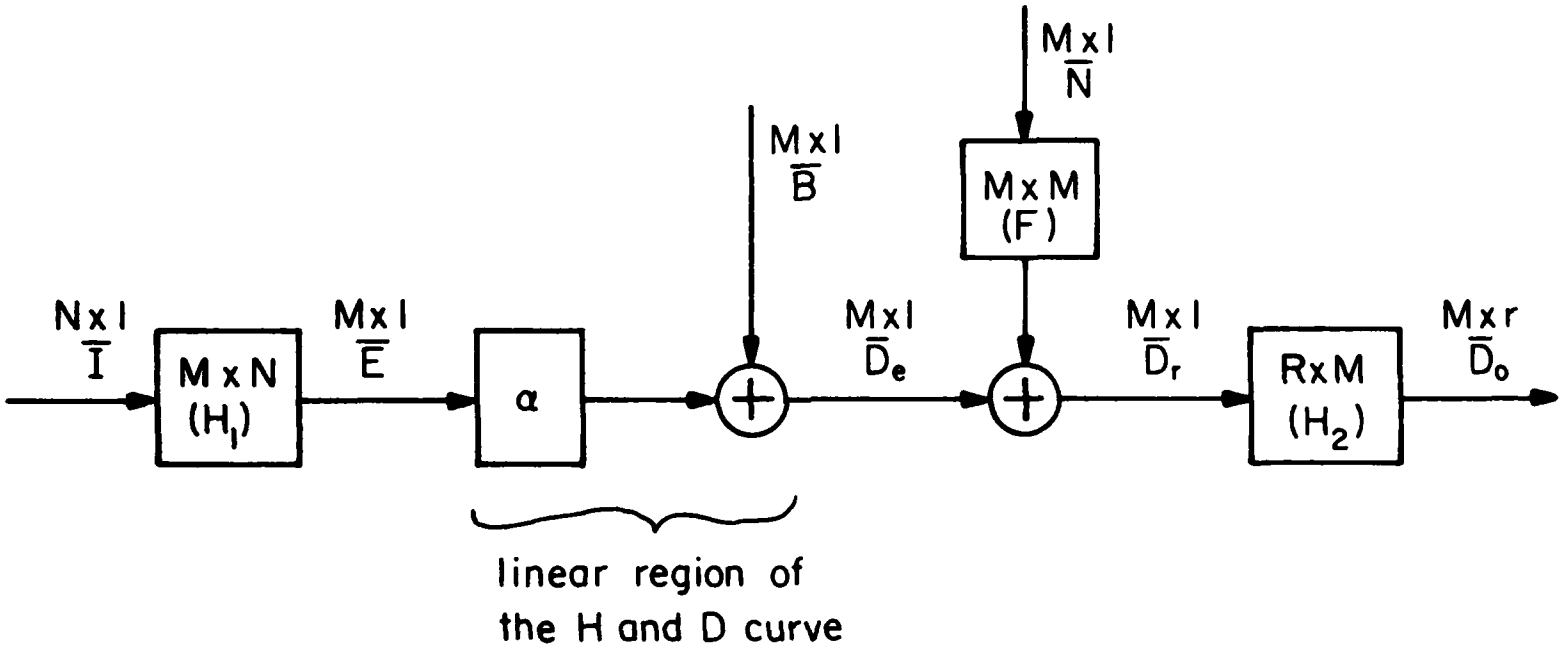


Figure 3.2-1. The discrete model.

operator. Note that matrix $[F]$ is structured to conform with the scalar eq. (3.2-1). The $R \times M$ matrix H_2 represents the combined effects of linear blurs $b_3(x, y)$ and $b_4(x, y)$ in the continuous model. Finally vector $\bar{D}_o$ is the observed degraded signal. The restoration filter should provide an estimate for the input signal $\bar{I}$ based on observation $\bar{D}_o$ and some a priori statistical information. Denoting this filter by an $N \times R$ matrix $[W]$, the estimate will be given by

$$\hat{\bar{I}} = [W]\bar{D}_o \quad . \tag{3.2-5}$$

It has been shown [6] that the optimal filter has the form

$$[W] = [C_{\bar{I}\bar{D}_o}][C_{\bar{D}_o\bar{D}_o}]^{-1} \tag{3.2-6}$$

where the criterion of optimality is to minimize the expected value of the square of the estimation error given by

$$e = E(\bar{I} - \hat{\bar{I}})^T(\bar{I} - \hat{\bar{I}}) \quad . \tag{3.2-7}$$

Here $(C_{\bar{I}\bar{D}_o})$ and $(C_{\bar{D}_o\bar{D}_o})$ are the cross-covariance of the signal and the observation and autocovariance of the observation respectively. It should be noted that the filter $[W]$ of eq. (3.2-6) is a linear non-causal batch processor. No assumption of stationarity is imposed in the derivation of this filter, and the general form as given by eq. (3.2-6) is independent of any underlying observation model. However, the evaluation of the two covariance matrices requires the knowledge of an observation model. We now devote the rest of this section to

the evaluation of this filter according to the discrete model of Fig. 3.2-1.

A. Evaluation of $[C_{\bar{I}\bar{D}_o}]$

Using the definition of cross covariance we have

$$[C_{\bar{I}\bar{D}_o}] = E(\bar{I}\,\bar{D}_o^T) - \bar{I}_m \bar{D}_{o_m}^T \tag{3.2-8}$$

where $\bar{I}_m$ and $\bar{D}_{o_m}$ are the nonstationary mean of the signal and the observation respectively. From Fig. 3.2-1 we can write

$$\bar{D}_o = [H_2](\alpha[H_1]\bar{I} + \bar{B} + [F]\bar{N}) . \tag{3.2-9}$$

Therefore

$$E(\bar{I}\,\bar{D}_o^T) = E\{\bar{I}(\alpha[H_1]\bar{I} + \bar{B} + [F]\bar{N})\cdot[H_2]^T\} . \tag{3.2-10}$$

Now let us assume that the signal $\bar{I}$ and noise $\bar{N}$ are independent, i.e.

$$f_{\bar{I}\bar{N}}(i_1, i_2, \ldots, i_N; n_1, n_2, \ldots, n_N) = f_{\bar{I}}(i_1, \ldots, i_N)\cdot f_{\bar{N}}(n_1, \ldots, n_N) \tag{3.2-11}$$

where $f_{\bar{I}\bar{N}}$ is the joint probability density function of $\bar{I}$ and $\bar{N}$. Using eq. (3.2-11) in eq. (3.2-10) we see that

$$E(\bar{I}\,\bar{D}_o^T) = \alpha E(\bar{I}\,\bar{I}^T)[H_1]^T[H_2]^T + \bar{I}_m \bar{B}^T[H_2]^T . \tag{3.2-12}$$

We now calculate the second term in eq. (3.2-8) namely $\bar{I}_m \cdot \bar{D}_{o_m}^T$. From eq. (3.2-9) we have

$$\bar{D}_{o_m} = [H_2](\alpha[H_1]\bar{I}_m + \bar{B}) \qquad (3.2\text{-}13)$$

and therefore

$$\bar{I}_m \bar{D}_{o_m}^T = \alpha \bar{I}_m \bar{I}_m^T [H_1]^T [H_2]^T + \bar{I}_m \bar{B}^T [H_2]^T. \qquad (3.2\text{-}14)$$

Using eqs. (3.2-12) and (3.2-14) in eq. (3.2-8) we obtain

$$[C_{\bar{I}\,\bar{D}_o}] = \alpha [C_{\bar{I}\,\bar{I}}][H_1]^T [H_2]^T \qquad (3.2\text{-}15)$$

where $[C_{\bar{I}\,\bar{I}}]$ is the autocovariance of the signal $\bar{I}$ given by

$$[C_{\bar{I}\,\bar{I}}] = E[\bar{I}\,\bar{I}^T] - \bar{I}_m \bar{I}_m^T. \qquad (3.2\text{-}16)$$

B. Evaluation of $[C_{\bar{D}_o\,\bar{D}_o}]$

Using the definition of autocovariance we have

$$[C_{\bar{D}_o\,\bar{D}_o}] = E(\bar{D}_o \bar{D}_o^T) - \bar{D}_{o_m} \cdot \bar{D}_{o_m}^T. \qquad (3.2\text{-}17)$$

From eq. (3.2-9) we have

$$E(\bar{D}_o \bar{D}_o^T) = E\{([H_2](\alpha[H_1]\bar{I} + \bar{B} + [F]\bar{N})) \cdot ([H_2](\alpha[H_1]\bar{I} + \bar{B} + [F]N))^T\}. \qquad (3.2\text{-}18)$$

Carrying out the above ensemble expectation and considering the independence of $\bar{I}$ and $\bar{N}$ we obtain

$$E(\bar{D}_o \bar{D}_o^T) = [H_2]\left(\alpha^2 [H_1] E(\bar{I}\,\bar{I}^T)[H_1]^T + \bar{B}\,\bar{B}^T + [C_{N'N'}] + \alpha[H_1]\bar{I}_m \bar{B}^T + \alpha \bar{B}\,\bar{I}_m^T [H_1]^T\right)[H_2]^T \qquad (3.2\text{-}19)$$

where

$$[C_{N'N'}] \triangleq E\{[F]\bar{N}\bar{N}^T[F]^T\} \tag{3.2-20}$$

in the above equation. Equation (3.2-2) and (3.2-4) indicate that $[C_{N'N'}]$ has the following form

$$[C_{N'N'}] = k^2 \begin{bmatrix} E(d_e^1)^{2P} & & & 0 \\ & E(d_e^2)^{2P} & & \\ & & \ddots & \\ 0 & & & E(d_e^M)^{2P} \end{bmatrix} \tag{3.2-21}$$

where d_e^j is the j^{th} element of vector $\bar{D}_e$ of Fig. 3.2-1. The second term in equation (3.2-17) is $\bar{D}_{o_m} \cdot \bar{D}_{o_m}^T$ and this can be calculated using eq. (3.2-9) as shown below

$$\bar{D}_{o_m} \cdot \bar{D}_{o_m}^T = [H_2]\left(\alpha^2[H_1]\bar{I}_m \cdot \bar{I}_m^T[H_1]^T + \bar{B}\bar{B}^T + \alpha[H_1]\bar{I}_m\bar{B}^T + \alpha\bar{B}\bar{I}_m^T[H_1]^T\right) \cdot [H_2]^T. \tag{3.2-22}$$

Finally using eq. (3.2-19) and (3.2-22) in eq. (3.2-17) we obtain

$$[C_{\bar{D}_o\bar{D}_o}] = [H_2]\left(\alpha^2[H_1][C_{\bar{I}\bar{I}}][H_1]^T + C_{N'N'}\right)[H_2]^T. \tag{3.2-23}$$

These expressions for $[C_{\bar{I}\bar{D}_o}]$ and $[C_{\bar{D}_o\bar{D}_o}]$ indicate that the filter $[W]$ has the form

$$[W] = \alpha[C_{\bar{I}\bar{I}}][H_1]^T\left(\alpha^2[H_1][C_{\bar{I}\bar{I}}][H_1]^T + [C_{N'N'}]\right)^{-1}[H_2]^{-1}. \tag{3.2-24}$$

The filter $[W]$ appears to have the form of a conventional discrete Wiener filter except for the presence of the blur matrix $[H_2]$, and more importantly, the structure of matrix $[C_{N'N'}]$. From Fig. 3.2-1 and eq. (3.2-21) we see that $[C_{N'N'}]$ is a diagonal matrix whose diagonal is given by

$$\text{Diagonal of } [C_{N'N'}] = \alpha k^2 \cdot E([H_1]\bar{I} + \bar{B})^{2P}. \qquad (3.2\text{-}25)$$

In Chapter 2 we discussed experimental results indicating that the exponent P in the above equation is less than $\frac{1}{2}$ and is more likely in the range 0.3-0.4. If for the sake of simplicity we take P to be $\frac{1}{2}$ we see that eq. (3.2-25) reduces nicely to

$$\text{Diagonal of } [C_{N'N'}] = \alpha k^2 [H_1]\bar{I}_m + \alpha k^2 \bar{B}. \qquad (3.2\text{-}26)$$

Equation (3.2-26) points out that the diagonal of the effective noise covariance matrix $[C_{N'N'}]$ is given by the scaled, blurred non-stationary a priori mean of the signal $\bar{I}$. It is exactly this feature that enables the filter to operate adaptively based on a priori information. Using the nonstationary expected value of the signal $\bar{I}_m$ to anticipate the level of the signal, the filter recognizes regions where the signal level is expected to be high as regions of high noise and filters accordingly. Conversely if the signal level is not expected to be high in a region, then the filter does not filter as much. This is the adaptive quality that we wanted to incorporate in the restoration filter.

In comparison, the diagonal elements of the covariance of the noise matrix in conventional discrete Wiener filter are all equal. Thus the filter assumes equal noise in all the regions, which as we have seen is not the case for images degraded by film-grain noise.

Before applying this filter to some one-dimensional and two-dimensional signals, we conclude this section by deriving an expression for the mean square error (mse) as given by

$$\text{mse} = \text{Trace}\{E[(\bar{I}-\hat{\bar{I}})(\bar{I}-\hat{\bar{I}})^T]\} . \qquad (3.2\text{-}27)$$

Using eq. (3.2-5) we have

$$\text{mse} = \text{Trace}\{E[(\bar{I}-[W]\bar{D}_o)(\bar{I}-[W]\bar{D}_o)^T]\} \qquad (3.2\text{-}28)$$

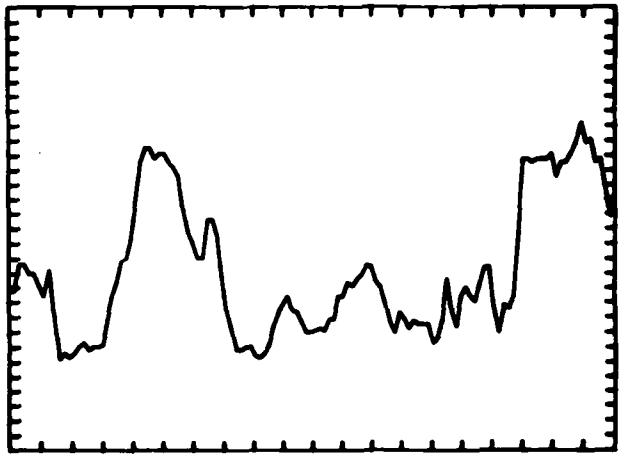
and carrying out the above expectation we get

$$\text{mse} = \text{Trace}\{[C_{\bar{I}\bar{I}}] - 2[W][C_{\bar{I}\bar{D}_o}]^T + [C_{\bar{I}\bar{D}_o}][W]\} \qquad (3.2\text{-}29)$$

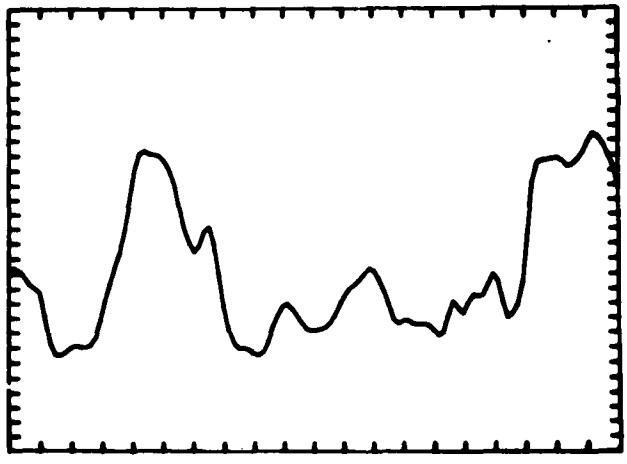
where all the terms in the above equation have previously been defined.

3.3 Experimental Results

As a preliminary evaluation of the performance of the filter derived in the last section we first applied it to a one-dimensional signal. Figure 3.2-2a shows the ideal signal corresponding to the input signal $\bar{I}$ of Fig. 3.2-1. This signal is obtained from one line of a typical digitized image, and the size of vector $\bar{I}$ is taken to be 128.



(a) Ideal signal

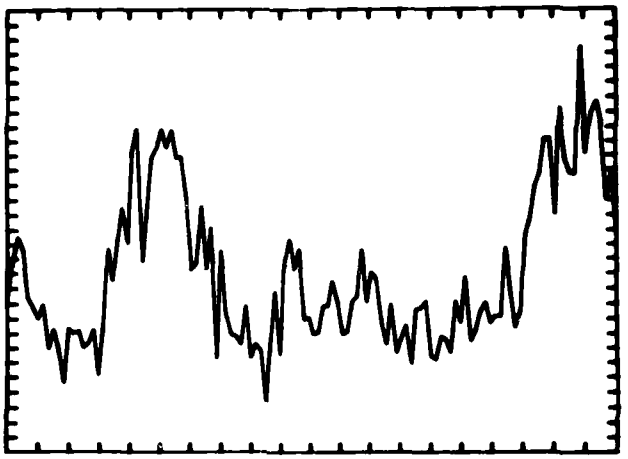


(b) Blurred signal

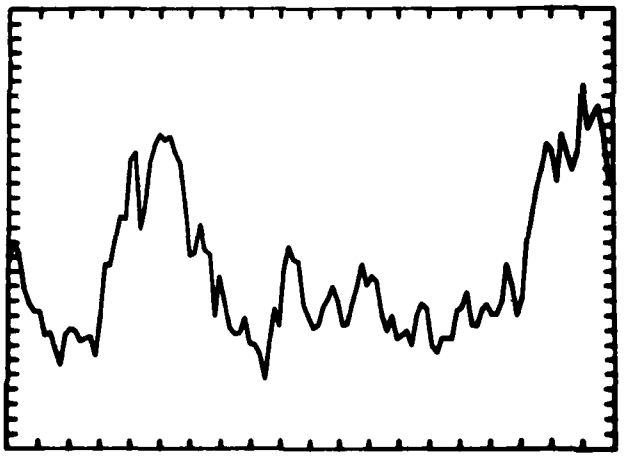
Figure 3.2-2. Simulation of degraded one dimensional signal.

We then degraded this signal in accordance with the model of Fig. 3.2-1. First, the vector $\bar{I}$ was blurred with matrix $[H_1]$ which was a Gaussian blur matrix with window of 5 elements and a σ of 0.9. The result is shown in Fig. 3.2-2b and corresponds to the vector $\bar{E}$ of Fig. 3.2-1. Because the constant α and constant vector $\bar{B}$ act as a scaling factors only, without any loss of generality, they were set to one and zero respectively. Therefore in this simulation, the vectors $\bar{E}$ and $\bar{D}_e$ of Fig. 3.2-1 are identical. A Gaussian signal-dependent noise was then added to vector $\bar{D}_e$ to obtain the signal $\bar{D}_r$ shown in Fig. 3.2-3a. The scanning constant k of eq. (3.2-2) was taken to be 0.3, corresponding to a fairly noisy film scanned with a very small aperture. Finally, vector $\bar{D}_r$ was blurred with matrix $[H_2]$ which was a Gaussian blur with a window of 3 elements and a σ of .5 to yield the observed signal $\bar{D}_o$ shown in Fig. 3.2-2b. This signal was processed by the LMMSE filter derived in the last section to yield an estimate $\hat{\bar{I}}$ for the input signal $\bar{I}$. The estimated signal is shown in Fig. 3.2-4b. Comparing figures (a) and (b) of Fig. 3.2-4 we note that the filter has filtered more in the high signal levels than the low signal levels. Overall the calculated mean-square-error was improved by a factor 2.5.

Note that in implementing filter $[W]$ of eq. (3.2-24) one needs to know $\bar{I}_m$ and $[C_{\bar{I}\bar{I}}]$, the first two moments of the signal. In this work $[C_{\bar{I}\bar{I}}]$ was approximated by the covariance matrix of a

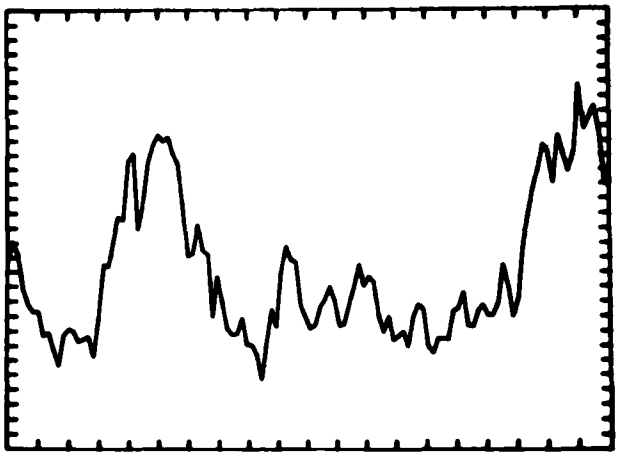


(a) Nonlinear noise added

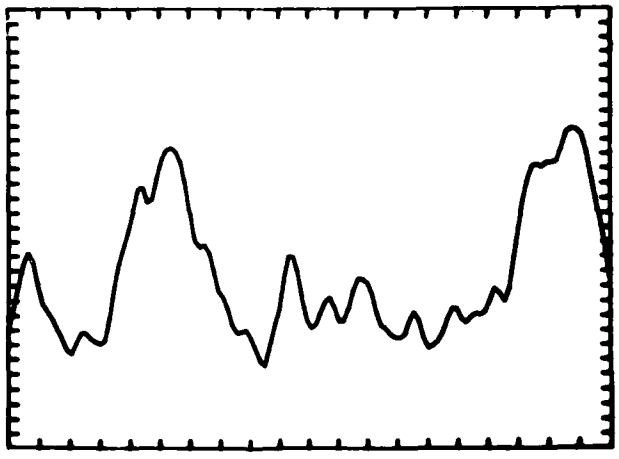


(b) Observed signal

Figure 3.2-3. Observed degraded signal.



(a) Observed signal



(b) Estimated signal

Figure 3.2-4. Restored one dimensional signal.

Markov process vector with first order correlation of 0.85. The a priori ensemble mean of the signal was estimated at each point by blurring the observed signal $\bar{D}_o$ with a very large window to obtain a rather smooth spatial average curve that was then used for $\bar{I}_m$. This procedure effectively approximates $\bar{I}_m$ by a local spatial average of $\bar{D}_o$.

For the restoration of two-dimensional signals or images one can arrange the pixels of the digitized image lexographically in a long vector and then proceed in the same manner discussed above for processing one-dimensional signals [8]. The problem with this procedure is that for a reasonable image of size 256 x 256 the resulting vector would have length 65,536. Obviously, the manipulation of vectors and matrices of this size becomes horrendous, even with large digital computers. An alternative method is to break up the image into small blocks and to process each sub-block individually. Similarly one can consider rows and columns of the image as separable one-dimensional signals and sequentially process the rows to obtain one estimate and then sequentially process the columns to obtain a second estimate. The final restored image is then a combination of these two estimates [6] - [9], with the simplest combination being the average. Note that both the block filtering or sequential row and column filtering result in suboptimal estimates.

In this section we present an example of two-dimensional

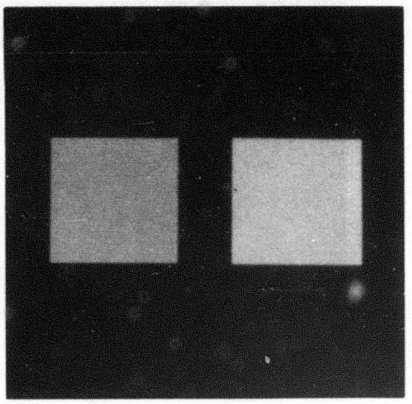
restoration using the sequential row and column processing approach. However before doing so an important point needs to be mentioned. Since we assume the image to have a nonstationary mean then the a priori mean for various lines of the image are not all equal. Equations (3.2-24) and (3.2-26) point out that every time the mean of a particular row differs from the mean of the previous row, a new covariance of the noise $[C_{N'N'}]$ must be calculated, in turn requiring a new filter $[W]$ to be calculated. Since any change in $[C_{N'N'}]$ necessitates performing a matrix inversion in order to calculate the filter $[W]$ of eq. (3.2-24), this causes an imposing problem. This problem is avoided by realizing that in most typical images of interest the means of adjacent lines do not differ substantially unless a sharp boundary is passed in going from one line to the other. Thus, the following algorithm is used in processing two-dimensional signals: When processing line j we compare the a priori mean for line j with the mean of the previous line; if no substantial difference exists we process line j with the same filter that was calculated for line j-1. If, however, the difference between the two means is judged to be considerable, then the new mean is used to calculate a new filter which is then used to process line j.

Figure 3.2-5 shows an example of the application of this filter to images. Figure (a) shows the ideal image. The three density levels are 0.4 for the background, 1.1 for the left hand square and

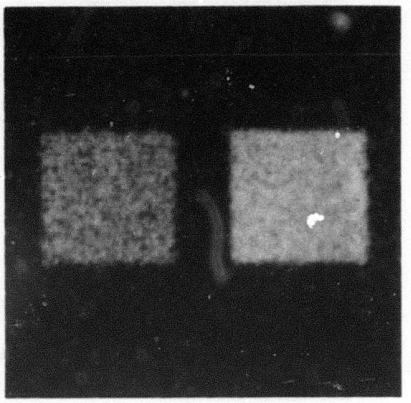
1.8 for the right hand square. The density for this image has a fairly large dynamic range; thus the noise in the degraded image is quite signal-dependent. Figure (b) shows the degraded image simulated according to the model of Fig. 3.2-1. In this simulation the initial blur is Gaussian with a window of 5 x 5 elements and σ of 0.9. The final blur is Gaussian with a window of 3 x 3 elements and σ of 0.5. The change in the a priori mean of the rows and columns requires the calculation of four different filters, each using a different mean for the matrix $[C_{N'N'}]$.

In Fig. 3.2-5 note that the filter has done more filtering in the two square areas where the noise is considerable. In the restored image there is some degree of smoothing. This smoothing is an inherent property of an estimator using mean-square error as the optimality criterion.

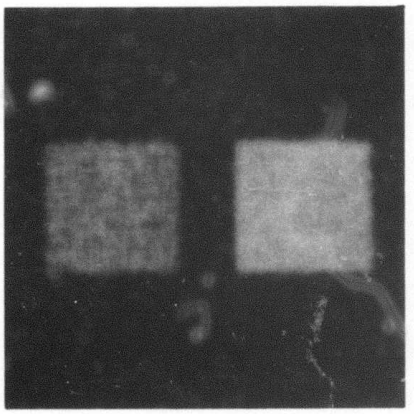
The performance of this filter can be compared with that of a conventional Wiener filter which ignores the signal-dependence of the noise. The result of the comparison is shown in Fig. 3.2-6. Comparing the square regions in the two restored images we see that the conventional Wiener filter does not filter these regions as well as the other filter because it assumes the noise to be the same in all the regions.



(a) Ideal image

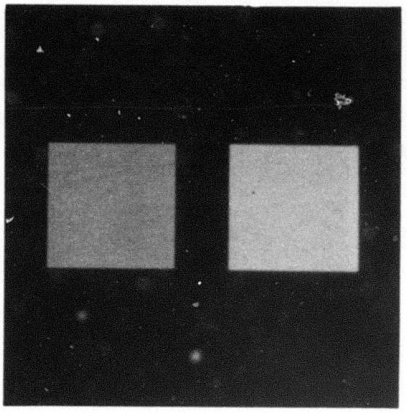


(b) Degraded image

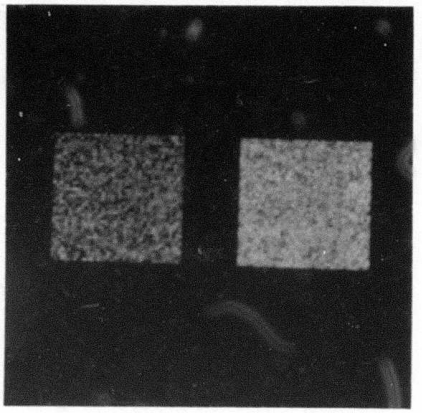


(c) Restored signal

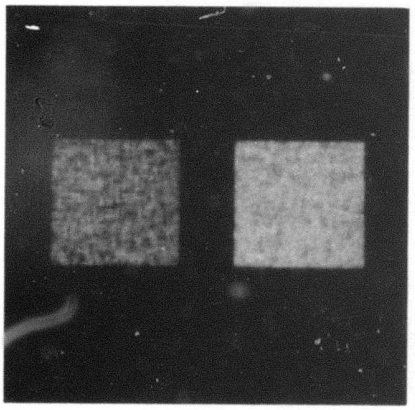
Figure 3.2-5. Two dimensional restoration.



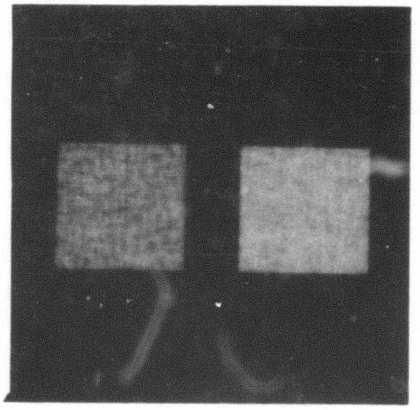
(a) Ideal image



(b) Degraded image



(c) Restored image by conventional Wiener filter



(d) Restored image by the filter of section 3.2

Figure 3.2-6. Comparison with conventional Wiener filter.

REFERENCES

[1] N. E. Nahi, Estimation Theory and Applications, J. Wiley & Sons, N.Y., 1969, Ch. 2.

[2] M. Naraghi, "An Algorithmic Image Estimation Method Applicable to Nonlinear Observations," University of Southern California Image Processing Institute Technical Report 580, University of Southern California, Los Angeles, California, 1975.

[3] B. R. Hunt, "Digital Image Processing," Proc. IEEE, Vol. 63, p. 693, 1975.

[4] B. R. Hunt and T. M. Cannon, "Nonstationary Assumptions for Gaussian Models of Images," to appear in IEEE Trans. Systems, Man and Cybernetics.

[5] J. F. Walkup and R. C. Choens, "Image Processing in Signal-Dependent Noise," Optical Engineering, Vol. 13, pp. 258-266, 1974.

[6] W. K. Pratt, "Generalized Wiener Filtering Computational Techniques," IEEE Trans. Computers, Vol. C-21, No. 7, pp. 636-641, July 1972.

[7] F. Davarian, "Fast Computational Techniques for Pseudo-Inverse and Wiener Image Restoration," University of Southern California Image Processing Institute Report 610, University of Southern California, Los Angeles, California, 90007, 1975.

[8] W. K. Pratt, Digital Image Processing, to be published.

[9] H. C. Andrews and B. R. Hunt, Digital Image Restoration, to be published.

Chapter 4

DETECTION I

In digitizing photographic films, it is sometimes necessary to use very small apertures in the scanning microdensitometer in order to detect fine details in the image. With these small apertures the film-grain noise is so severe that the detection of fine details and distinction between adjacent areas of small contrast becomes virtually impossible. The noise can be reduced by using a larger aperture for the scanner; however doing this will produce averaging over fine details of interest in the image. Thus in most cases, the photograph must be digitized using a small aperture and then improved by effective enhancement and restoration techniques.

When scanned with apertures as small as 4 to 25 $(\mu m)^2$, the digitized image is so degraded that conventional restoration techniques employing linear or nonlinear estimators will have little effect. In these cases heuristic but clever detection schemes taking the properties of the human observer into account appear to be more helpful than any estimation method. In this chapter we analyze an enhancement method called "noise cheating" introduced by H. J. Zweig et al. [1] and show that their work can be justified through classical detection theory. The "noise cheating" algorithm is an effective enhancement method for low contrast images having a small

dynamic range.

To expand on this algorithm and make it applicable to a broader class of images, and to develop new and more flexible algorithms of this type, we must provide a justification for the heuristic nature of the "noise cheating" algorithm. Without such justification, further expansion on the idea beyond the heuristic procedures becomes impractical. Thus, this chapter is devoted to analysis of "noise cheating" algorithm within the framework of detection theory. In the next chapter we improve this algorithm so that it is applicable to a broader class of images. We will also develop a Bayesian detection method which is considerably more flexible than the "noise cheating" algorithm.

The "noise cheating" method in essence functions as a two-stage detecting procedure. A detailed analysis of this two-stage procedure is discussed in the rest of this chapter.

4.1 Enhancement by Two-Stage Maximum Likelihood Detection

Assume that in the absence of grain-noise, the section of photograph to be scanned is composed of M extended regions $R_1, \ldots, R_M$ all of them of much larger size than the scanning aperture and each of them having a uniform density. Denote by $D_1, \ldots, D_M$ the density levels of these M regions and assume the image is extremely low contrast so that values of $D_1, \ldots, D_M$ are very close.

Suppose this film is scanned with a small square aperture of

size $2a \times 2a$. If a is on the order of 1 - 3 μm., then the M regions will no longer be distinguishable from one another in the digital image due to severity of the noise and the low contrast. Thus some sort of enhancement is necessary.

To begin our analysis of the "noise cheating" algorithm, we start with a set of simplifying assumptions. Though unrealistic, these assumptions will facilitate the understanding of the nature of the algorithm and once this understanding is reached these assumptions will be dropped.

The first assumption is that we know a priori that the image consists of exactly M distinct regions, even though we cannot make out all of these M regions in the degraded image. We also know the density levels of these regions; thus we assume a priori knowledge of values for $D_1, \ldots, D_M$.

Based on each single reading of the scanner (each single reading of the scanner corresponds to a pixel in the digitized image) we want to decide if the reading was taken when the densitometer aperture was completely in one of the M regions or if it was taken with the aperture overlapping two or more of the regions. If this decision can be made, then by changing the value of each observed pixel to the density of the detected region, some of the noise effects should be suppressed.

To achieve this, we formulate the problem as the following M+1

hypothesis detection problem. Corresponding to each single reading produced by the scanner, we denote by $H_1, H_2, \ldots, H_M$ the hypothesis that the reading was taken with the scanner's aperture completely in regions $R_1, R_2, \ldots, R_M$ respectively. Let H_{M+1} denote the hypothesis that the reading was taken with the aperture overlapped by two or more regions. Figure 4.1-1 illustrates these hypotheses.

As we discussed in Chapter 2, the output of the scanner is a function of the mean density of the region scanned and the granularity noise. Denoting the scanner's output by y, we write

$$y = g(D, n) \tag{4.1-1}$$

for the observation.

Detection of hypotheses $H_1, \ldots, H_{M+1}$ by a maximum likelihood detection scheme with y as the observation is equivalent to announcing hypothesis H_k if

$$f(y|H_k) > f(y|H_i) \qquad \begin{matrix} i = 1, \ldots, M+1 \\ i \neq k \end{matrix} \tag{4.1-2}$$

where $f(y|H_i)$ is the conditional probability density of y under hypothesis H_i [2]. Therefore, the evaluation of these M+1 conditional probability densities are required for detection.

Guided by the results of Chapter 2, for the first M hypothesis, the conditional density can be readily written

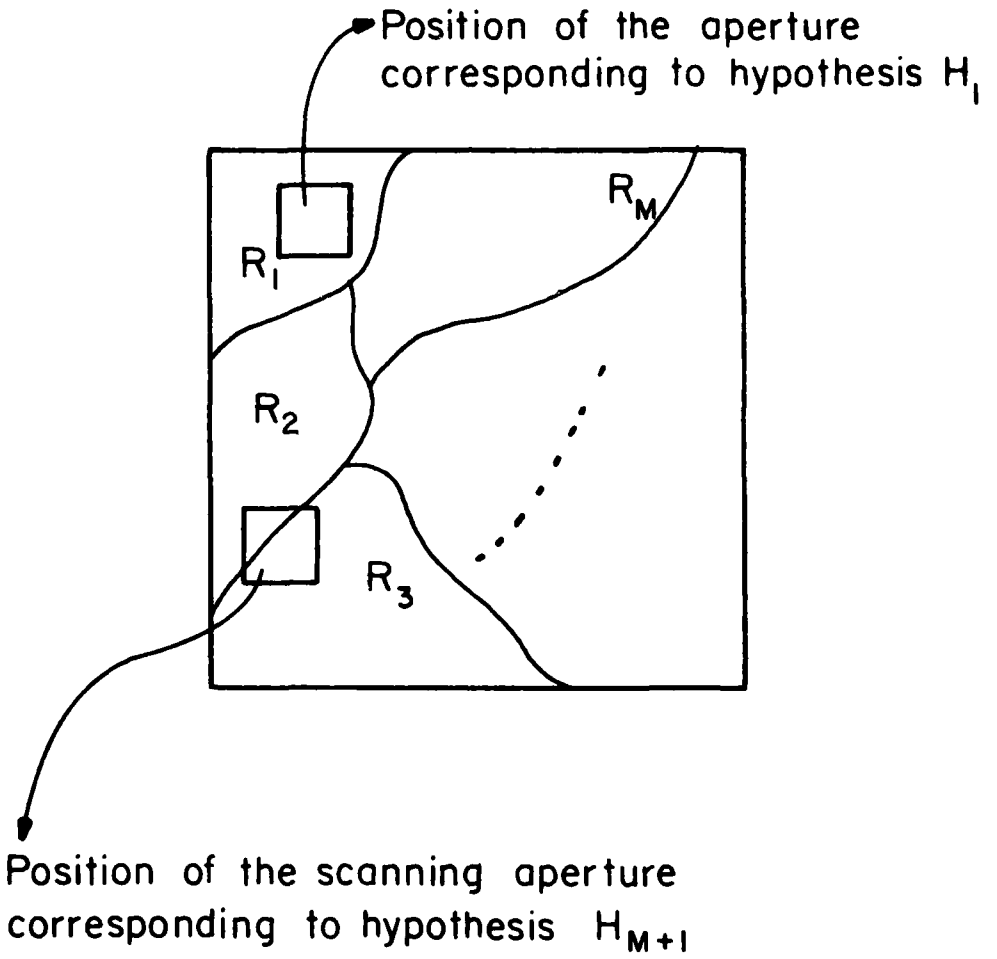


Figure 4.1-1. Position of the scanning aperture under various hypothesis.

$$f(y|H_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \; e^{-[(y-D_i)^2/2\sigma_i^2]} \qquad i = 1, \ldots, M \tag{4.1-3}$$

where

$$\sigma_i = kD_i^{1/3} \; . \tag{4.1-4}$$

However $f(y|H_{M+1})$ is far more complicated. In general we can write

$$f(y|H_{M+1}) = \frac{1}{\sqrt{2\pi}\sigma_{M+1}} \; e^{-\{[y-(\alpha_1 D_1 + \alpha_2 D_2 + \ldots + \alpha_M D_M)]^2/2\sigma_{M+1}^2\}} \tag{4.1-5}$$

where $0 \leq \alpha_i < 1$. The exact form of this conditional density depends on α_i's which in turn depends on the number of regions and the amount of overlap in each as shown in Fig. 4.1-1. It is clear that no unique expression for $f(y|H_{M+1})$ can be written because there are an infinite number of expressions for this conditional density depending on different combinations of α_i's. In fact, the set containing all the possible expressions for this density is uncountable. Therefore, carrying out the maximum likelihood detection test as indicated by eq. (4.1-2) becomes impossible.

To avoid this problem, let us consider a suboptimal detection procedure. We perform the detection in two stages. In the first stage we ignore the overlap hypothesis H_{M+1} and assume that all the readings were taken with the aperture completely in one of the M regions. We then perform M hypothesis maximum likelihood

detection on the image based only on hypotheses $H_1, \ldots, H_M$.

In the second stage we examine the detected image and in areas where hypothesis H_{M+1} appears very likely to have occurred (i.e. the boundary of the M regions) we refine the details of the image according to another algorithm.

The net effect is that we accurately define the coarse features of the image, i.e. the rough boundaries of the M extended regions, in the first stage and sharpen the edges in the second stage.

4.2 Detection by Quantization

In this section we consider the first stage of the algorithm just discussed and see how it can be implemented by a simple quantization.

In the previous section we made the assumption that there is a small difference between the density levels of different regions. We can therefore rewrite eq. (4.1-3) as

$$f(y \mid H_i) = \frac{1}{\sqrt{2\pi}\sigma} \; e^{-\left[(y-D_i)^2/2\sigma^2\right]} \quad i = 1, \ldots, M \qquad (4.2\text{-}1)$$

where σ_i has been replaced by σ. We can choose σ to be

$$\sigma = k(D)^{1/3} \qquad (4.2\text{-}2)$$

where D may arbitrarily be taken as the mean value of the D_i's. Thus we have used the low contrast of the image to omit the dependence of σ on the D_i's and thus in effect neglect the signal-dependence of the noise.

Without loss of generality, we assume that $D_1 < D_2 < \ldots < D_M$. It can now easily be shown that the maximum likelihood detection rule of selecting H_k if

$$f(y|H_k) > f(y|H_i) \quad \begin{matrix} i = 1, \ldots, M \\ i \neq k \end{matrix} \tag{4.2-3}$$

is the same as selecting H_k if and only if

$$L_k < y < L_{k+1} \tag{4.2-4}$$

where L_k and L_{k+1} are given by

$$L_k = \frac{D_k + D_{k-1}}{2} \tag{4.2-5}$$

$$L_{k+1} = \frac{D_k + D_{k+1}}{2} \quad . \tag{4.2-6}$$

These equivalent statements become clear by a simple inspection of Fig. 4.2-1. Equations (4.2-3) to (4.2-6) also point out an interesting fact. If we quantize each pixel of the degraded image with a quantizer whose quantization levels are placed at density levels D_i, $i = 1, \ldots, M$ and whose decision levels are placed at half way between the density levels at

$$L_i = \frac{D_i + D_{i-1}}{2} \quad i = 2, \ldots, M \, , \tag{4.2-7}$$

then in effect we perform the maximum likelihood detection as described by eq. (4.2-3). Therefore the first stage of detection can be implemented by quantization of the degraded image with the quantizer described above.

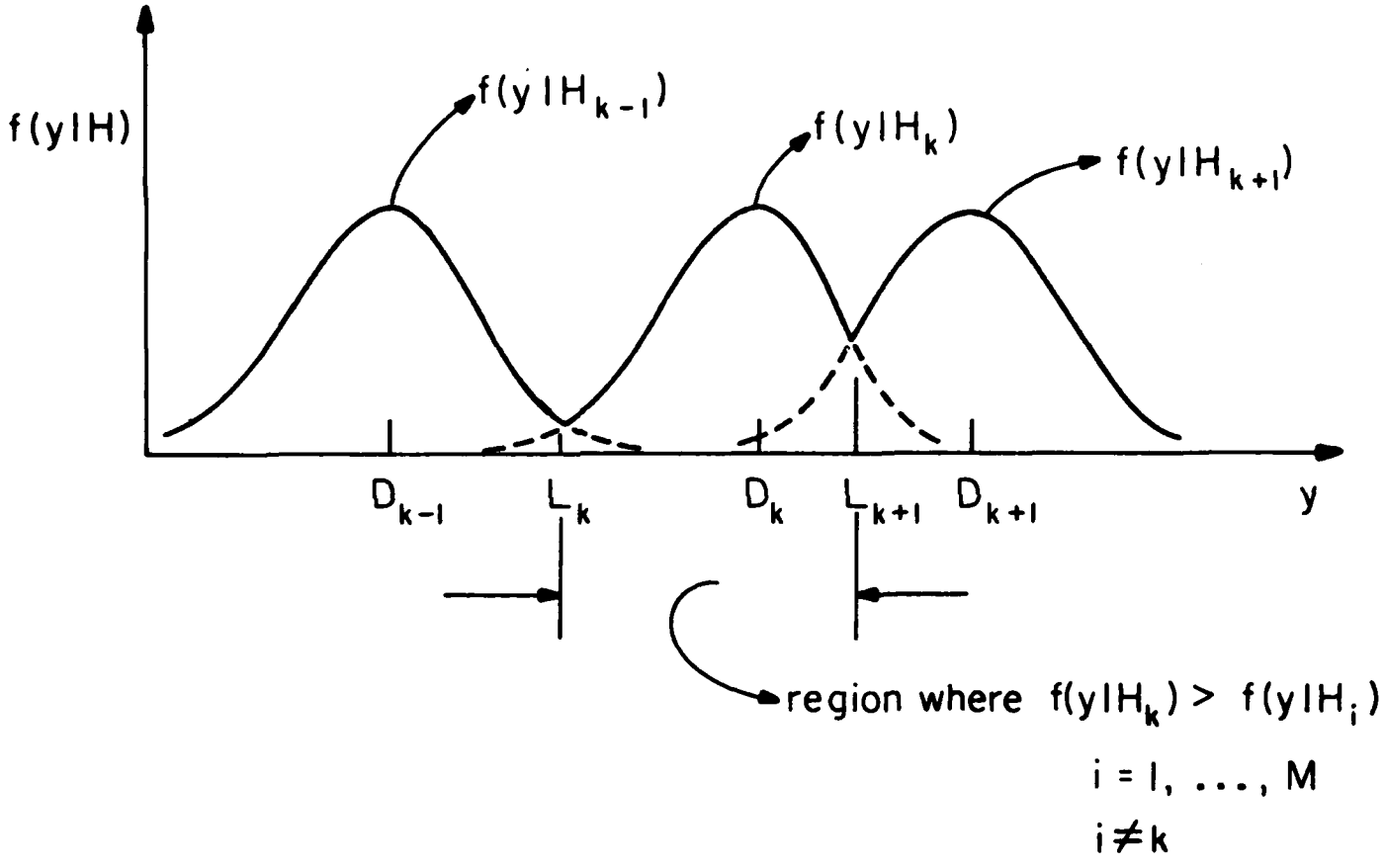


Figure 4.2-1. Conditional probability density function of the scanner output under different hypotheses.

Let us now evaluate the performance of the quantizer by examining the quantized image at the end of the first stage. Suppose in digitizing the image we select the $2a \times 2a$ aperture of the scanner such that the resulting noise standard deviation as given by

$$\sigma = \frac{k'}{\sqrt{2a \times 2a}} (D)^{1/3} \qquad (4.2\text{-}8)$$

is no more than 1/4 of the difference between the two closest D_i's. Therefore all D_i's would be at least 4σ apart. Selection of the aperture under the above condition would thus imply that the decision levels of the quantizer used in the first stage would be at least 2σ to each side of their corresponding quantization levels as seen in Fig. 4.2-2.

Since pixels in the degraded image belonging to region R_i are Gaussian distributed with mean D_i and standard deviation σ, then roughly 95% of the readings taken by the densitometer from region R_i will be in a band 2σ to each side of D_i and therefore will be subsequently quantized to D_i.

Due to the properties of Gaussian distribution [3], this means that after the first stage of detection, 95% of the interior points of each region will be correctly detected and restored.

4.3 Probability of Correct Detection in the First Stage

In the previous section we saw that with a probability of 0.95 the interior points of uniform density regions are correctly detected and

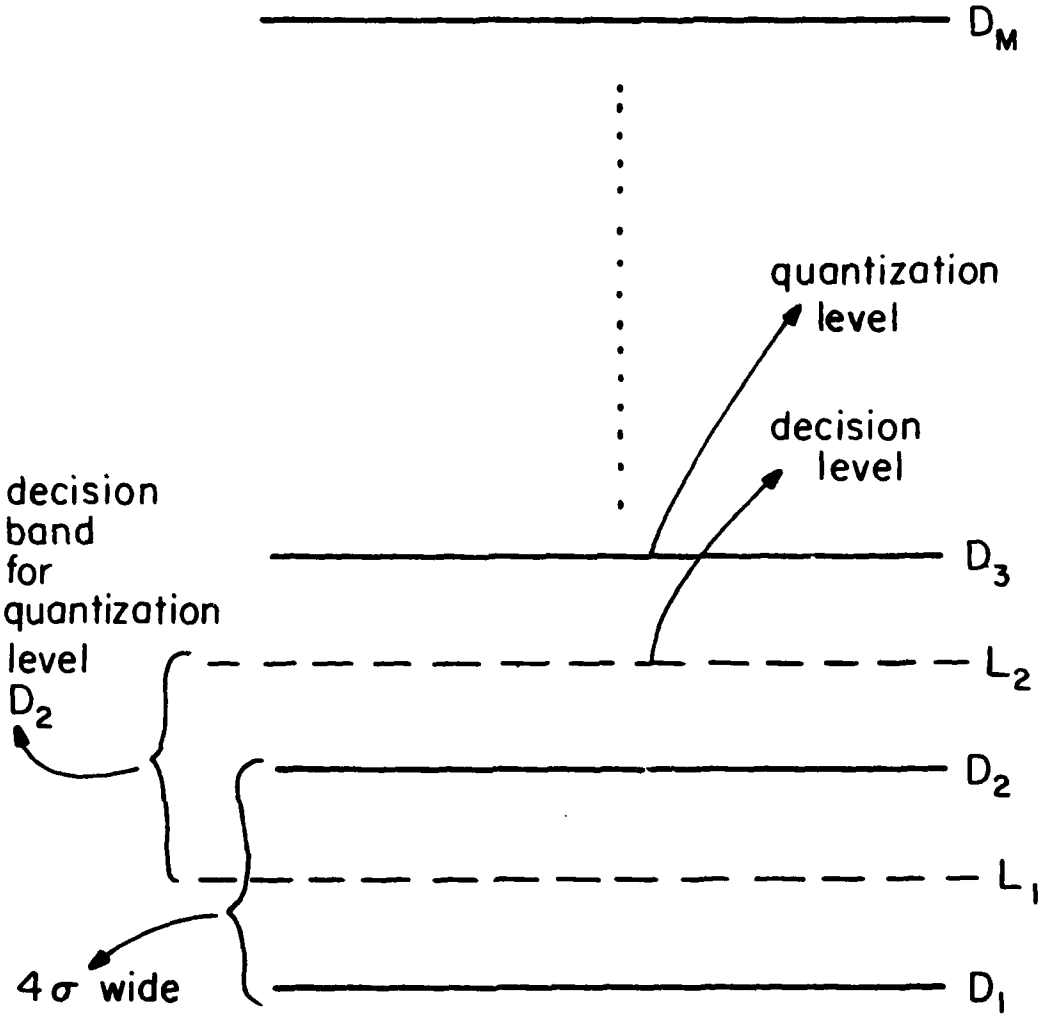


Figure 4.2-2. The quantizer used in the first stage of the detection.

restored to the right density. These results are conditioned on the selection of an aperture for the scanner so that the noise standard deviation σ is no more than 1/4 of the difference between two closest D_i's.

From eq. (4.2-8) we see that the standard deviation of the noise is inversely proportional to the scanning aperture. Suppose we digitize the image with a larger aperture such that the resulting noise standard deviation is only 1/6 of the difference between the two closest D_i's. This means that the decision bands of the quantizer now extend to at least 3σ to each side of the quantization levels. Due to the Gaussian distribution of the scanner output, the probability of correct detection for the interior points of each region increases to 0.97 with this larger aperture.

From the above argument it is clear that as the aperture size is increased, the distance between the quantization levels increases (in terms of σ) and the probability of correct detection increases for the interior points. There is a tradeoff, however, because the large aperture will average over a large area to produce a single reading, thus destroying some of the fine detail at the boundaries of the M regions.

A smaller aperture preserves more information about the edges, but also decreases the probability of correct detection for the interior points. To illustrate this point, suppose we use a smaller

aperture resulting in a noise σ which is 1/2 the difference between the two closest D_i's. Again following the same line of argument, we see that the decision bands of the quantizer used in the first stage are now only 2σ wide or one σ to each side of the quantization levels. Therefore the probability of correct detection for the interior points is only 0.67.

A compromise can be reached by using two different size apertures. A large aperture of size $2a \times 2a$ can be used in the first stage of the detection. The size of this aperture will be selected so that the distance between any two D_i's is at least 4σ and thus the probability of correct detection for the interior points is a high 0.95.

As we said before, the second stage of the detection is designed to compensate for the omission of hypothesis H_{M+1} in the first stage. Recall that this situation occurs when the aperture overlaps two or more of the extended regions. Clearly this hypothesis is the most probable when the scanner is digitizing areas around the boundary of the M extended regions.

Since we have detected the interior points of each region at the end of the first stage with low probability of error, we have roughly outlined the boundaries of the extended regions where hypothesis H_{M+1} is most likely to have occurred. The second stage of the detection should thus involve a closer examination of the regions which have been designated to be boundary areas by the first stage.

In the next section we show how these areas may be redigitized with a smaller aperture to improve resolution.

4.4 Second Stage of the Detection

After the first stage of the detection and prior to the second stage involving boundary operations, all isolated points are removed from the detected image. An isolated point is defined to be a pixel having a value which is different from all of its closest eight neighboring pixels. In removing the isolated pixel the value of the isolated pixel is changed to conform with that of its neighbors. This is justified because we assume the picture to consist only of extended regions several times the size of the scanning aperture. Thus we can conclude that any isolated pixel was erroneously detected.

After the first stage we define a pixel to belong to a boundary if one or more of its eight closest neighboring pixels have been detected to have a value different from it. Since hypothesis H_{M+1} is most likely to have been true at the boundaries, all the boundary points need to be reexamined.

The reexamination is achieved by redigitizing the photograph with a smaller aperture at the areas which have been designated to be boundaries by the first stage. This new data is used to improve the resolution.

Suppose the smaller aperture is of size a x a and thus is only 1/4 of the size of the aperture originally used. This would mean that

for each single pixel produced by the scanner during the first digitization, we now obtain four readings.* We can now use the newly obtained data to sharpen the boundary. An algorithm used by Zweig et al. [1] is explained below by an example.

Referring to Fig. 4.4-1a and allowing each small square to be a pixel in the digitized image, suppose that after the first stage we have roughly outlined four regions having densities 0.6, 0.75, 0.8, and 1.1. According to our definition, the center pixel with value 0.8 is a boundary pixel and has to be reexamined. By redigitizing with a smaller aperture of 1/4 of the size originally used, we obtain 4 new readings corresponding to this one pixel as shown in Fig. 4.4-1b. We replace the pixel with these newly obtained data as shown in Fig. 4.4-1c. Now once more we have to decide to which one of the regions these new pixels should be assigned. At this point an interesting point can be raised. Equation (4.2-8) points out that the new data obtained by redigitizing the photograph is twice as noisy as

*In practice there is no need to physically scan the image twice, once with each aperture. It suffices to scan the image with the smaller of the two apertures. Then the computer can be used to perform a 2 x 2 averaging on the digital image to effectively obtain what we would have obtained had we actually scanned the image a second time with the larger aperture.

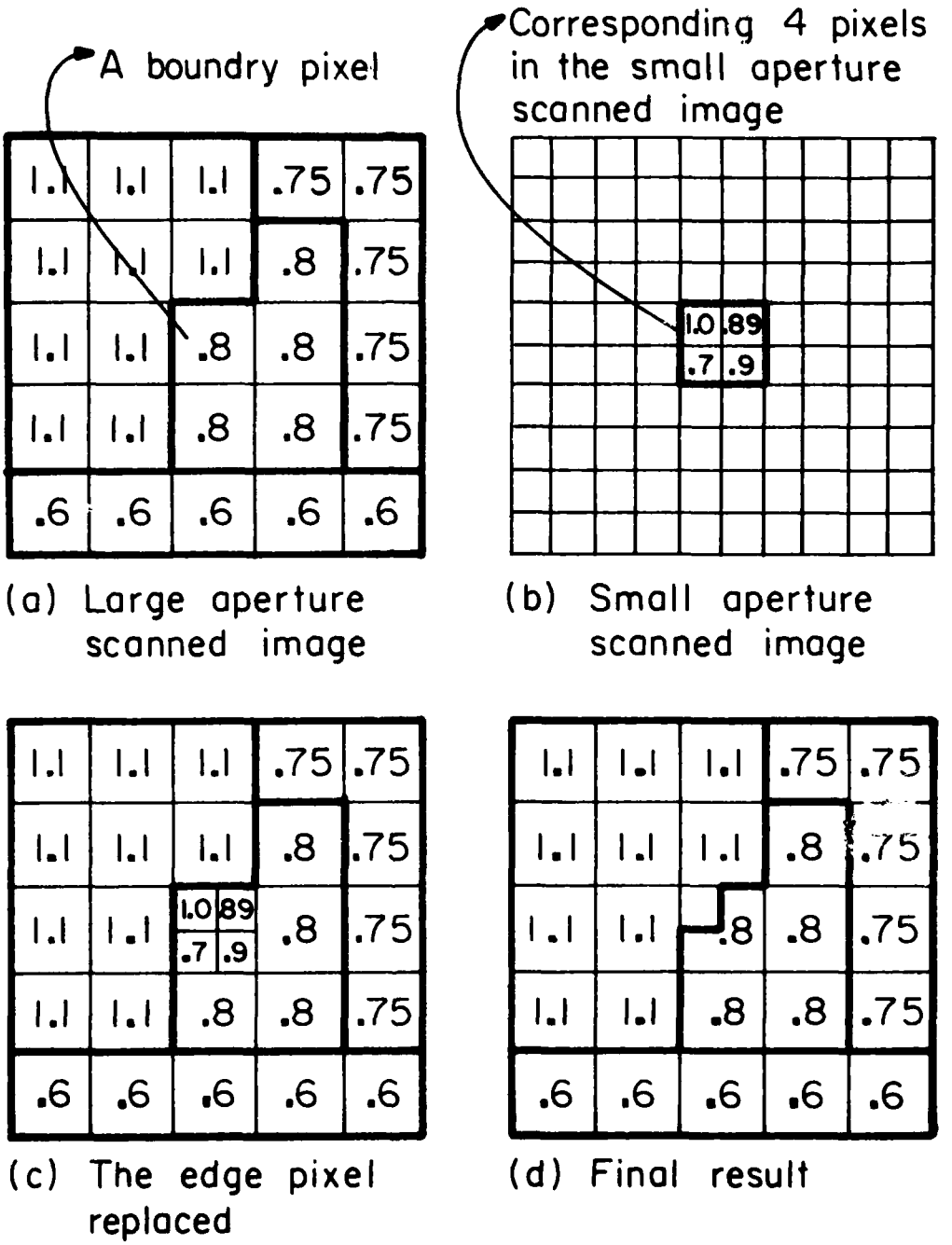


Figure 4.4-1. Second stage of the detection.

the data which was used originally in connection with the first stage of the detection. This occurs because the scanning aperture is reduced in size by 1/4. Therefore a logical question is to ask how the noisier data can be used to sharpen the boundaries detected using data less noisy.

One possible answer is now described. In the first stage, after observing each pixel in the degraded digitized image, we had to decide to which one of M regions the pixel should be assigned. In the second stage, even though we are using noisier data, we no longer need to consider all of the M regions in order so assign the small pixels to the correct value. We need only consider the immediate vicinity. This point can be made clear by the example of Fig. 4.4-1. In Fig. 4.4-1c there are four new observations having values 1.0, 0.89, 0.7 and 0.9. Note that after the first stage we have already made a determination that the area these 4 pixels occupy lies at the boundary of the two extended regions having values 1.1 and 0.8. Therefore with these new four data points we have only to decide to which one of the two regions, 1.1 or 0.8, we should assign the new pixels. This is in contrast to the first stage where we decide between M regions in assigning each pixel. In other words, in the second stage we do conditional detection in the sense that in reassigning each boundary point we consider only the regions

on whose boundary the point lies.

The actual reassignment can be done according to different criteria. One possible choice is according to the absolute value of the difference. For example in Fig. 4.4-1c of the 4 new data points, pixels having 0.89, 0.7 and 0.9 are assigned to the region having density 0.8 while the fourth pixel having value 1.0 is assigned to the region density 1.1. Comparing Fig. 4.4-1a and 4.4-1d we see how the data from the second digitization of the photograph is used to modify and sharpen the boundaries between regions having densities 0.8 and 1.1. Similarly all other boundary points should be reworked.

This finishes explanation of the basic structure of the algorithm. However, if the complete algorithm is to be useful, in practice, some of the unrealistic assumptions that were made in the derivation must now be dropped. This is done in the next section.

4.5 Detection with Unknown Signal Levels

In the previous analysis we assume that the degraded image consisted of exactly M regions with uniform densities, even though the regions could not be identified exactly. We further assumed that the density levels of these regions were known a priori. Clearly these assumptions are not realistic. In this section we intent to omit these assumptions and reexamine the detection algorithm discussed earlier in the chapter.

We will retain the basic two stage structure of the detection

algorithm. In the first stage we use a quantizer to implement the detector and use as observations the data obtained by scanning the film with an aperture of size 2a x 2a. The second stage is not altered because all the assumptions that we are going to relax affect the first stage only. Recall that the quantization levels of the quantizer used in the first stage were set at density levels D_i, $i = 1, \ldots, M$ which were assumed known. Since we no longer are going to make this assumption, we need to restructure the quantizer to be used in the first stage.

As before, we assume that the images are composed of several extended regions each having a uniform density. We assume that the number of these regions and their density levels is not known, however we do assume that the image has very low contrast. We examine low contrast images because they are particularly vulnerable to film-grain noise.

To construct a quantizer for use in the first stage, we begin by arbitrarily choosing a quantization level Q_0. Having set this quantization level, we place other quantization levels above and below Q_0 such that they are exactly four noise standard deviations apart. Since the image has low contrast, one can approximate the standard deviation of the noise by

$$\sigma = \frac{k_2}{\sqrt{2a \times 2a}} D^{1/3} \qquad (4.5\text{-}1)$$

where the density D can be taken to be the sample mean taken over the entire degraded digitized image.

Let the quantization levels thus obtained be denoted by

$$Q_L, \ldots, Q_{-2}, Q_{-1}, Q_0, Q_1, Q_2, \ldots, Q_U$$

where the difference between each successive pair is 4σ. The actual number of quantization levels depends on the expected dynamic range of the image density. As discussed in Chapter 2, for almost all classes of images the density will vary in the range of 0.0 to 2.5, therefore we arbitrarily select Q_0 somewhere in this range and then place other quantization levels 4σ apart until this range is filled up.

Having set the quantization levels, we set the decision levels at the mid-point between each pair of quantization levels. Thus all the decision bands are 4σ wide. We have now constructed a quantizer without any a priori knowledge.

Now we investigate the consequence of utilizing this quantizer in the first stage of the detection algorithm discussed earlier and following it with the second stage which is left unchanged.

As an example, suppose we digitize a small section of a photograph by scanning it with a very small square aperture of size 2a×2a. The digitized image will be degraded with a film-grain noise having standard deviation σ as given by eq. (4.5-1). Suppose the section

of the photograph that we have scanned is composed of four extended regions R_1, R_2, R_3 and R_4 with unknown boundaries and unknown uniform densities D_1, D_2, D_3 and D_4. If the image has low contrast, these four regions will be indistinguishable from one another in the digitized image due to the small size of the scanning aperture and enhancement is necessary.

We now apply the two stage detection algorithm using the quantizer discussed in this section in the first stage. To fully understand the nature of this algorithm let us assume that in constructing the quantizer, it just happens that D_1, the density level of region R_1, coincides exactly with quantization level Q_k. Assume also that D_2, the density level of region R_2, happens to be one σ greater than quantization level Q_k and that D_3 and D_4 happen to be 2σ and 3σ greater than Q_k respectively, as shown in Fig. 4.5-1.

Using this quantizer, all pixels in the degraded image having values between L_k and L_{k-1} will be quantized to Q_k. On the other hand we know that pixels in the degraded image belonging to region R_1 are Gaussian distributed with mean D_1 and standard deviation σ. Therefore, since D_1 coincides with Q_k and since L_k and L_{k-1} are 2σ to each side of Q_k, then upon quantizing the degraded image 95% of pixels belonging to region R_1 will be quantized to D_1 while 2.5% will be quantized to Q_{k+1} and another 2.5% will be quantized to Q_{k-1}.

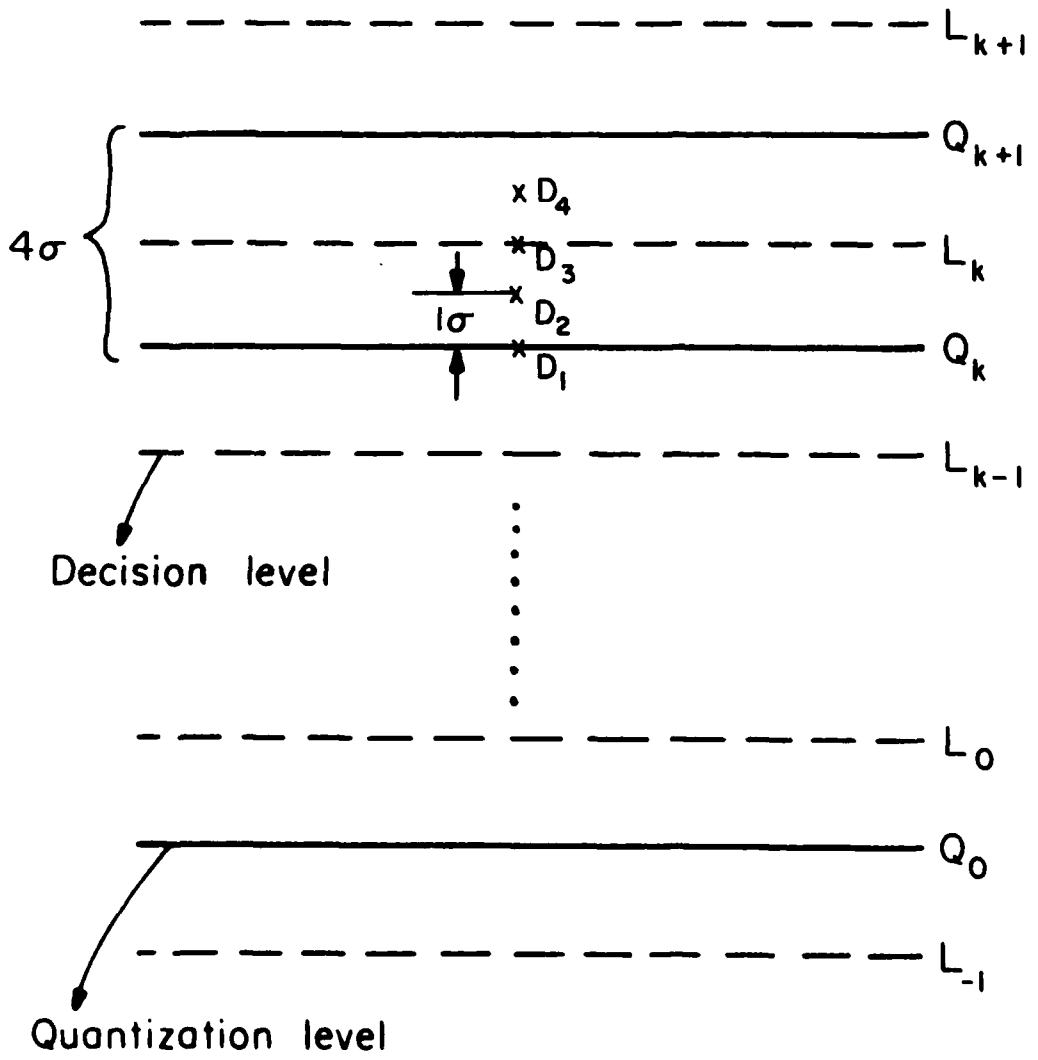


Figure 4.5-1. Structure of the quantizer when standard deviation of the noise is σ.

Thus after the first stage only 5% of the pixels belonging to region R_1 are erroneously assigned to the wrong density.

As outlined before, we remove all isolated points after quantization and prior to the second stage. For a pixel in region R_1 to be an isolated point after the first stage, it should be an erroneously detected pixel surrounded by 8 pixels which have correctly been detected to have value D_1. Thus the probability of a pixel being an isolated point is $(.95)^8$ or roughly 0.66. Thus, by removing all isolated points from this region, of the original 5% falsely detected pixels, 0.66x5% = 3.3% are corrected to have value D_1 leaving only 1.7% misclassified pixels in the interior of this region prior to second stage of the detection.

In the second stage of processing the edge pixels are reworked. Because 1.7% of the pixels have values other than D_1 within the interior of region R_1, then all these pixels and their 8 closest neighbors will qualify as edge pixels. Therefore 9 x 1.7 or about 15% of the pixels in region R_1 will be redetected using data obtained by scanning the photograph a second time with smaller aperture size axa. Of the 15% reworked pixels some will ultimately be redetected to have value D_1 thus bringing the probability of correct detection in region R_1 to well over 0.85.

The above completes analysis of the performance of the detection

algorithm for regions having a density which just happens to coincide with a quantization level of the quantizer used in the first stage of the algorithm. However, since the quantization levels are placed quite arbitrarily, it is not very likely that many regions in the image will have a density that will coincide exactly with a quantization level. Thus it is more interesting to consider the fate of regions R_2, R_3 and R_4 which have densities that fall between two quantization levels rather than on any particular one.

We first consider region R_2. This region has a density level D_2 which exceeds quantization level Q_k by one σ and is less than quantization Q_{k+1} by 3σ. Since degraded digitized image pixels belonging to region R_2 are Gaussian distributed about mean D_2 with standard deviation σ, and since decision levels L_{k-1} and L_k of the quantizer are 3σ and one σ away from D_2 respectively, then upon quantization, 84% of the pixels in R_2 will be quantized to Q_k while 16% will be quantized to Q_{k+1}.

After quantization a pixel, quantized to Q_{k+1} has a $(.84)^8$ or .25 probability of being an isolated point. Meanwhile a pixel quantized to Q_k has $(.16)^8$ or about zero probability of being an isolated point. Thus in the process of removing isolated points, .25 x 16% or about 4% of pixels originally quantized to Q_{k+1} will be changed to Q_k. Therefore, prior to the second stage of the detection, 88% of the interior points of region R_2 are given value Q_k while the remaining

12% are given value Q_{k+1}.

In reexamining the boundary pixels in the second stage of the detection, the 12% pixels having value Q_{k+1} and their 8 closest neighboring pixels will be considered boundary pixels. Thus $9 \times 12\% = 98\%$ or roughly the entire region R_2 will be reexamined in the second stage.

Recall that in the second stage we use data obtained by scanning the film with a smaller aperture of size a x a to rework the boundary pixels. Clearly in region R_2 this new set of data will once again be Gaussian distributed with mean D_2 but now with standard deviation σ' which is twice as large as σ from eq. (4.5-1).

Since pixels in region R_2 after the first stage of detection have values Q_k or Q_{k+1}, the newly obtained data will be assigned value Q_k or Q_{k+1} depending on the absolute value of the difference between the observed data and these two values as explained in section 4.4. Referring to Fig. 4.5-1 we note that in the new set of data, pixels with a value greater than L_k will be closer to Q_{k+1} while pixels with a value less than L_k will be closer to Q_k. Now since L_k is $\sigma'/2$ away from D_2, from Gaussian distribution properties it is clear that 31% of pixels in the new data will be greater than L_k and thus will be quantized to Q_{k+1} while the other 69% will be assigned value Q_k. Thus in the final detected image, region R_2 will be composed of pixels of which 69% have value Q_k and 31% have value

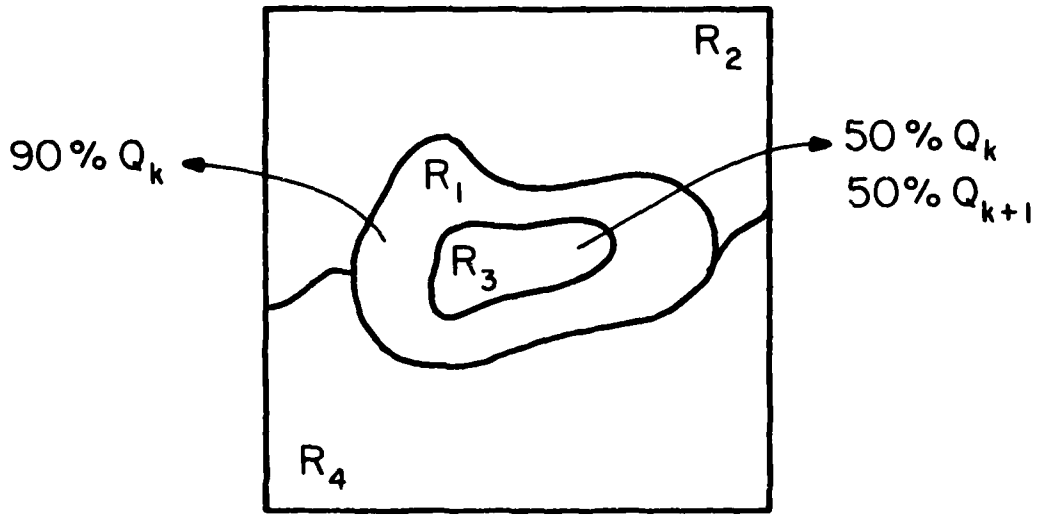
Q_{k+1}.

Following the same line of argument we used to analyze regions R_1 and R_2, one can show that in the final detected image region R_3 will consist of pixels having value Q_k 50% of the time and value Q_{k+1} the remaining 50%. Region R_4 will then be composed 69% of pixels with value Q_{k+1} and 31% with value Q_k. Thus various regions will be coded into various percentages of quantization levels.

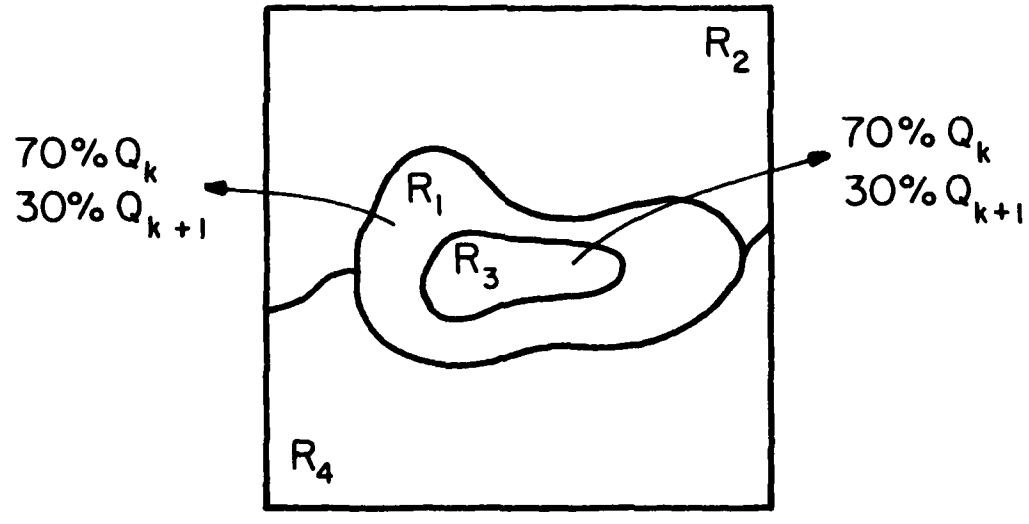
4.6 Effectiveness of the Algorithm for Enhancement

At first one might question the usefulness of this detection algorithm. In the case just discussed, we detect in region R_3 about 50% of the pixels to have value Q_k and the other 50% to have value Q_{k+1}, while in reality all the pixels in this region have a value exactly half the distance between Q_k and Q_{k+1}. Thus the probability of error in this region is 1. However in image detection involving the human observer the primary goal is distinction between regions of different density levels rather than the density value itself. Therefore the conventional detection error probability is not indicative of the performance of this detector.

To see that this algorithm accomplishes the goal of image detection, consider the case just discussed in which region R_3 is surrounded by region R_1 as shown in Fig. 4.6-1a. Since the image pixels in region R_1 are detected to have value Q_k about 90% of the time while pixels in region R_3 are detected to have value Q_k 50% of the



The case when D_1 falls on Q_k



The case where D_1 falls one σ below Q_k

Figure 4.6-1. Composition of regions in two different cases.

time and value Q_{k+1} the other 50%, the two regions will be easily distinguishable from one another in the detected image. This is due to the fact that human eye can easily separate regions having such a significant difference in structure. The experimental results of the next chapter will bear this out.

Referring to Fig. 4.5-1 we see that the density levels of regions R_1 and R_3 are 2σ apart. To better understand the algorithm, suppose we construct our quantizer so that D_1 happens to fall one σ below quantization level Q_k rather than on Q_k as assumed earlier, and that D_3 falls one σ above Q_k. In the final detected image we will find that region R_1 is detected as having value Q_k 69% of the time and value Q_{k-1} 31% of the time, while in region R_3 the percentages are 69% Q_k and 31% Q_{k+1}. In this case both regions have value Q_k 69% of the time and the eye will not be able to distinguish these two regions as easily as before because their structures are not sufficiently different.

The structures of these and all other regions in the final detected image can be changed by shifting the quantization levels by a small amount. In section 4.4 we initially selected quantization level Q_0 arbitrarily and then placed other levels such that adjacent pairs were 4σ apart. As discussed in the example of Fig. 4.6-1b the mean densities of various regions relative to these quantization levels in the image might fall in locations so that some adjacent regions would

get coded into a very similar percentage of quantization levels and thus not be sufficiently distinguishable from one another in the final detected image.

If we shift the original quantization level Q_0 by one σ this means that all quantization levels are shifted by one σ and the location of mean densities of various regions changes relative to these new quantization levels. Since it is the relative distance between the quantization levels and the mean densities that determines how regions are coded into various percentages of the quantization levels, then the appearance of the final detected image varies substantially as the quantization levels are shifted. One can repeatedly shift the quantization levels by small intervals until the human observer is satisfied with the final result.

In the discussion above we have allowed the human observer to be in a feedback loop in order to improve the final results. Naturally there are some cases where image restoration techniques have to be designed so that they are capable of processing hundreds of degraded images in a very short time. In such cases one obviously cannot insert the human observer in the feedback loop because of insufficient time for individual inspection of each restored image. However such volume processing is seldom encountered when the primary goal of restoration is the elimination of film-grain noise.

Before concluding this chapter we should give some insight into

selection of the aperture for the scanner. The selection of appropriate size aperture is of utmost importance and there are two main factors to consider. The most obvious one is the size of the smallest object one expects to find in the section of the film being scanned. If one is looking for objects of size 40 $(\mu m.)^2$ the size of the scanner aperture must be considerably smaller than 40 $(m\mu.)^2$ more likely on the order of 4 $(\mu m.)^2$. This is done so that the scanner will not average over a large portion of the object.

The second factor to be considered is the contrast of the adjacent regions. If we hope to distinguish two adjacent regions whose densities are different by ΔD, we should select our aperture such that the resulting noise standard deviation is no more than $\Delta D/2$ to insure that the density levels of these two regions are at least 2σ apart and therefore upon quantization they are not often assigned to the same quantization level in the final detected image.

REFERENCES

[1] H. J. Zweig, et al., "Noise-Cheating Image Enhancement, J. Opt. Soc. Am., Vol. 65, pp. 1347, Nov. 1975.

[2] J. M. Wozencraft, I. M. Jacobs, Principle of Communication Engineering, John Wiley & Sons, N.Y., 1967, pp. 211-214.

[3] A. Papoulis. Probability, Random Variables, and Stochastic Processes. McGraw-Hill, N.Y., 1965, p. 64.

Chapter 5

DETECTION II

In Chapter 4 we used the fundamentals of detection theory to provide a framework to analyze the "noise cheating" algorithm. In this chapter we use the same framework to develop more versatile algorithms.

In the last chapter we considered a class of low contrast images with small dynamic range. The assumption of small range for the optical density enabled us to omit the signal-dependence of the noise and take σ, the standard deviation of the noise to be invariant over the entire image. Many images have a large dynamic range in density, yet have regions of local low contrast. In such cases, ignoring the signal-dependence of the noise affects the performance of any detection scheme adversely.

We start this chapter by modifying the detection scheme discussed in the last chapter to obtain a detector which takes into account the signal-dependence and thus is applicable to a broader class of images. In the second half of this chapter we develop a more versatile Bayesian detection scheme enabling us to accommodate any a priori knowledge available.

5.1 A Detector for Signal-Dependent Noise

The need for considering signal-dependence in film-grain noise restoration of images with dynamic range, becomes clear by examining the signal-to-noise ratio (SNR) of the degraged image.

In a local portion of an image when the mean density is D we can define signal to noise ratio to be

$$\mathrm{SNR}(D) = \frac{D^2}{\sigma_D^2} \tag{5.1-1}$$

where σ_D is standard deviation of the noise. We equivalently can write

$$\mathrm{SNR}(D) = \frac{D^2}{(kD^{1/3})^2} = \frac{1}{k^2} D^{4/3} . \tag{5.1-2}$$

Let us consider a Kodak Pan-X film which has been scanned with a 4 $(\mu m.)^2$ aperture. The scanning constant k for this film and this size aperture is roughly 0.6. Now suppose that there are two regions in the digitized image, one having mean density of 0.2 and the other having mean density of 2.2. Calculating the SNR for these two regions, from eq. (5.1-2) we have

$$\mathrm{SNR}(.2) = \frac{1}{(.6)^2} (.2)^{4/3} = 0.319 \tag{5.1-3}$$

and

$$\mathrm{SNR}(2.2) = \frac{1}{(.6)^2} (2.2)^{4/3} = 8.478 . \tag{5.1-4}$$

Thus in this case SNR for the two regions of this image varies by a factor of 26 and there is a definite advantage in considering the signal-dependence in processing this degraded image.

With the above motivation, we now describe modifications to the "noise cheating" algorithm discussed in the last chapter in order to incorporate the signal-dependence effects. In section 4.4 we outlined a procedure for the construction of a quantizer utilized in the first stage of the two-stage detection scheme. In constructing this quantizer we place the initial quantization level Q_0 arbitrarily. The other quantization levels are then uniformly spaced at 4σ intervals when σ, the standard deviation of the noise, is assumed to be a constant for all the various regions in the image. In quantizing the degraded image with this quantizer we can make the following remarks:

a) If a region in the degraded image has a density whose value coincides with a particular quantization level, then roughly 90% of the pixels in this region are correctly detected and their values changed to the density of the regions.

b) If a region in the degraded image has a density whose value falls in between two quantization levels, then after quantization all the pixels in this region are assigned a value from one of the two quantization levels.

Both of the above statements depend on the condition that all decision

levels for the quantizer are 2σ away from their quantization levels.

Equation (2.3-10) points out that σ, the standard deviation of the noise, is a monotonically increasing function of the mean density. Thus if we insist on placing the quantization levels of the quantizer 4σ apart, consistently, then after arbitrarily selecting the initial quantization level the other levels should be spaced successively wider as we move toward high density values and successively smaller as we move toward zero density. In the last chapter we were able to place the quantization levels at uniform spacing because we assumed that the images of interest had a very small dynamic range and thus σ varied very little from one region of the image to another. If a quantizer with uniformly spaced quantization levels is used to detect an image with a large dynamic range, the effects will be:

a) Adjacent regions in the degraded image having very small densities that are not substantially different will be detected as the same region because the distance between the quantization levels is far more than the deviation of the pixels in these regions from their mean densities.

b) Regions with high densities will be erroneously quantized to several levels because in the degraded image some pixels in these regions deviate from the density of the region by more than the distance between the decision levels of the quantizer. This will

cause a region of a uniform density to appear in the detected image as a region having several densities.

In order to avoid these effects we must make sure that the quantization levels of the quantizer are consistently 4σ apart, where σ is allowed to increase with the mean density according to eq. (2.3-10). Accordingly we decide on the following procedure to construct the quantizer. We place the initial quantization level arbitrarily at Q_0. The next quantization level greater than Q_0 is denoted Q_1 and is found such that

$$Q_1 - Q_0 = 2\sigma_{Q_0} + 2\sigma_{Q_1} \tag{5.1-7}$$

where

$$\sigma_{Q_0} = k(Q_0)^{1/3} \tag{5.1-8}$$

and

$$\sigma_{Q_1} = k(Q_1)^{1/3} \quad . \tag{5.1-9}$$

Combining eq. (5.1-7) with eqs. (5.1-8) and (5.1-9) we get

$$Q_1 - 2k(Q_1)^{1/3} - Q_0 - 2k(Q_0)^{1/3} = 0 \tag{5.1-10}$$

or equivalently

$$Q_1^3 - 3\alpha Q_1^2 + (3\alpha^2 + 8k^3)Q_1 - \alpha^3 = 0 \tag{5.1-11}$$

where

$$\alpha = Q_0 + 2\sigma_{Q_0} \quad . \tag{5.1-12}$$

Since Q_0 is known, then α can be calculated from eqs. (5.1-8) and (5.1-12). Evaluation of Q_1 thus involves solution of the third order equation given by (5.1-11). This equation has a deterministic solution [1] of the form

$$Q_1 = A+B+\alpha \tag{5.1-13}$$

where

$$A = \sqrt[3]{\frac{-q}{2} + \sqrt{R}} \tag{5.1-14}$$

and

$$B = \sqrt[3]{\frac{-q}{2} - \sqrt{R}} \tag{5.1-15}$$

and where

$$R = \left(\frac{P}{3}\right)^3 + \left(\frac{q}{2}\right)^2 \tag{5.1-16}$$

$$P = 6\alpha^2 - 8k^2 \tag{5.1-17}$$

$$q = -8\alpha k^2 \ . \tag{5.1-18}$$

After Q_1 is found, we repeat this procedure to find Q_2, Q_3 and so on until a quantization level exceeds the upper bound of the expected dynamic range for the optical density of the image. In most cases, this is less than 2.5.

We can also follow the same procedure to find quantization levels less than Q_0. For example, Q_{-1}, the first quantization level smaller than Q_0 can be calculated

$$Q_0 - Q_{-1} = 2\sigma_{Q_0} - 2\sigma_{Q_1} \ . \tag{5.1-19}$$

We continue finding levels Q_{-2}, Q_{-3}, etc. until we get a quantization level which is smaller than zero. Because the standard deviation as calculated from

$$\sigma_{Q_i} = k(Q_i)^{1/3} \qquad (5.1\text{-}20)$$

becomes extremely small as Q_i approaches zero, there will be an infinite number of closely packed quantization levels near zoer. To avoid this, we may decide a priori that below a certain level, say 0.3, we abandon this procedure and set all the quantization levels at a uniform spacing of $4\sigma_i$, where σ_i is obtained from eq. (5.1-20) and Q_i is the first quantization level greater than or equal to 0.3.

Figure 5.1-1 shows a set of quantization levels obtained according to this procedure with the scanning constant k = .12 and the initial quantization level $Q_0 = 0.4$.

As discussed in section 2 of chapter 4, we would like to find a set of decision levels for this quantizer corresponding to maximum likelihood detection of regions whose mean densities happen to coincide with one of the quantization levels. Consider decision level L_k which is to be placed in between quantization levels Q_k and Q_{k+1}. Any two regions in the degraded image having mean densities equal to Q_k and Q_{k+1} will have pixels which are Gaussian distributed about mean Q_k and Q_{k+1} with standard deviations σ_{Q_k} and $\sigma_{Q_{k+1}}$ as given by eq. (5.1-20). The graphs of these two Gaussian probability

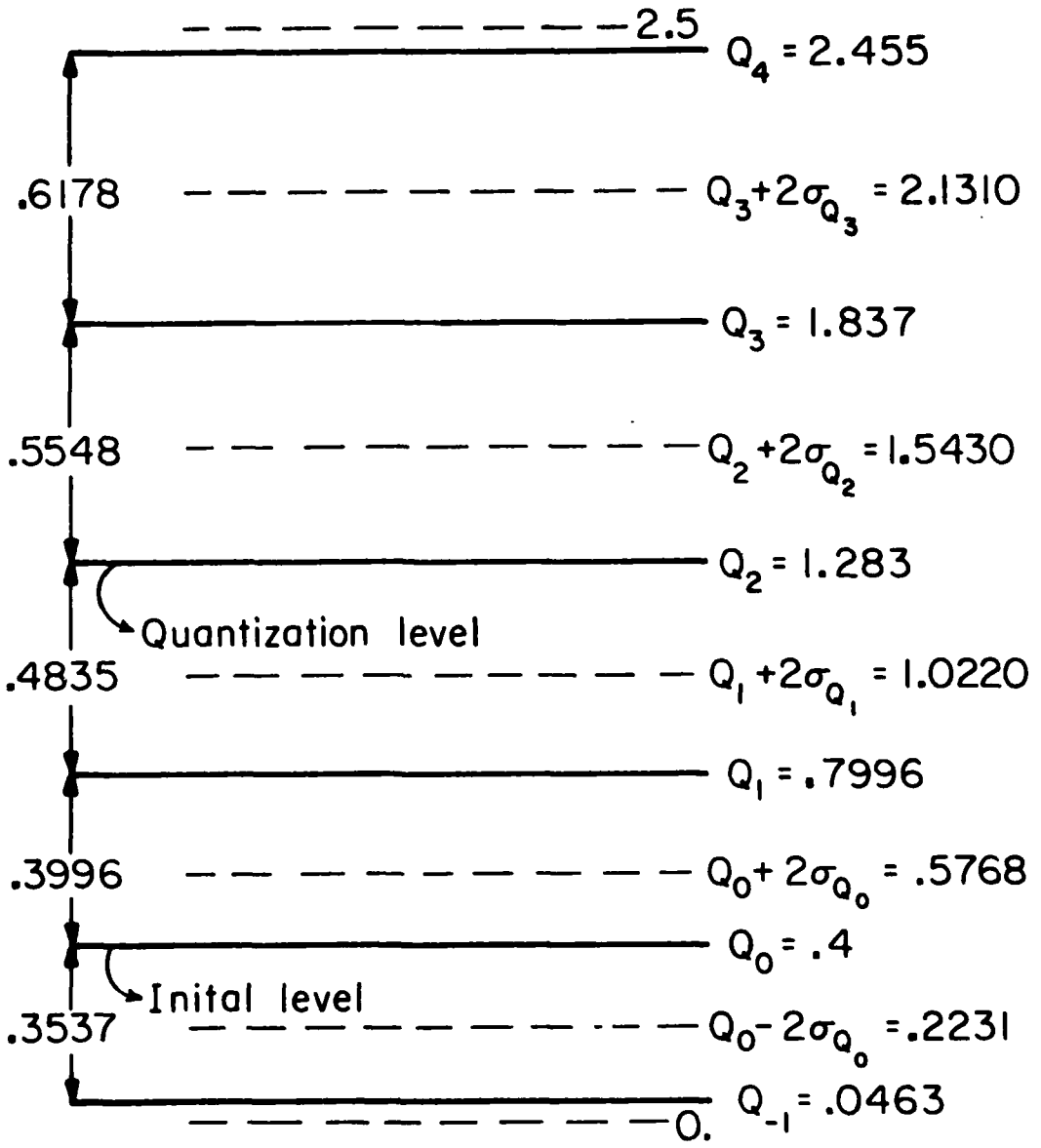


Figure 5.1-1. An example of nonuniformly spaced quantization levels.

distribution functions will intersect one another at a point h where

$$h = \frac{\sigma^2_{Q_{k+1}} \cdot Q_k - \sigma^2_{Q_k} \cdot Q_{k+1} \pm \sigma_{Q_k} \sigma_{Q_{k+1}} \cdot \Delta}{\sigma^2_{Q_{k+1}} - \sigma^2_{Q_k}} \qquad (5.1\text{-}21)$$

and where

$$\Delta = \sqrt{(Q_{k+1}-Q_k)^2 - 2 \ln\left(\frac{\sigma_{Q_k}}{\sigma_{Q_{k+1}}}\right)\left(\sigma^2_{Q_{k+1}} - \sigma^2_{Q_k}\right)}. \qquad (5.1\text{-}22)$$

Note that Λ is always positive, so that h will always have two real solutions of which only one lies between Q_k and Q_{k+1}. A Gaussian probability density with mean Q_{k+1} and variance $\sigma^2_{Q_{k+1}}$ lies above a Gaussian distribution with mean Q_k and variance $\sigma^2_{Q_k}$ for all points greater than h and will lie below this function for all points less than h. Accordingly, we set the decision level L_k at h so that quantization with this quantizer corresponds with maximum likelihood detection in the sense described earlier. Figure 5.1-2 shows a pair of quantization levels and their corresponding decision level obtained according to the above procedure.

The two-stage detection algorithm now involves quantization of the degraded image with a quantizer constructed according to the procedure outlined in this section. As before, the rough boundaries of various regions in the image are detected. This is followed by a second stage of detection which is exactly the same as that described in section 4.4.

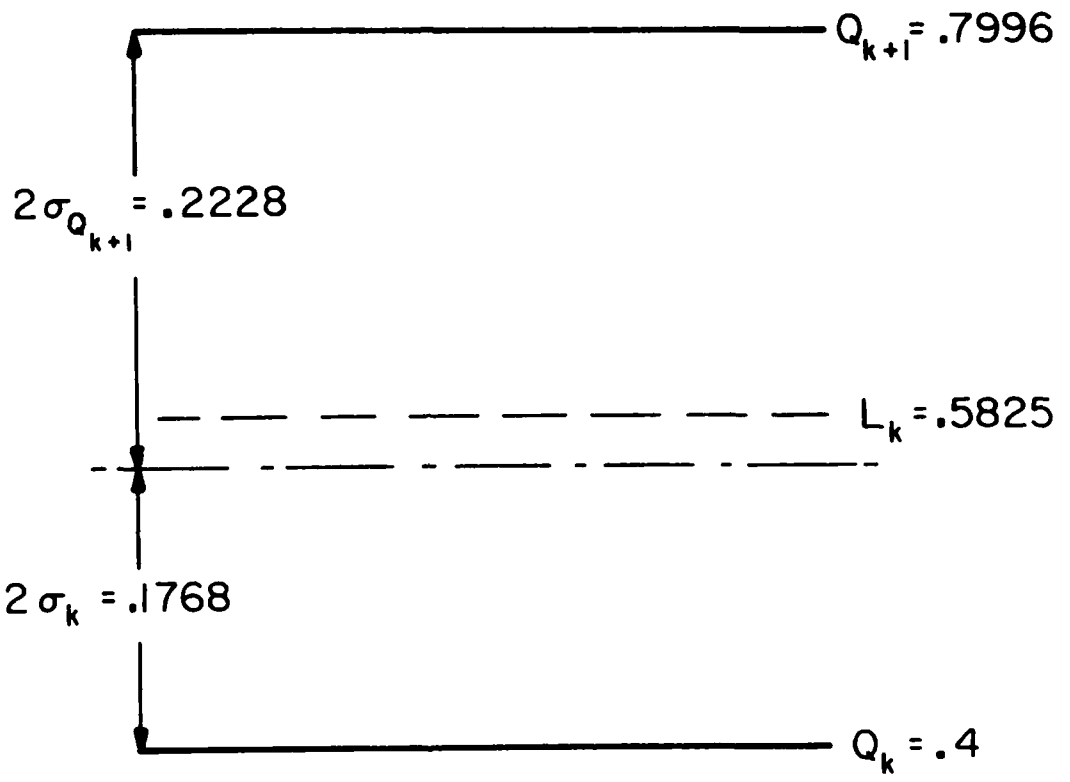


Figure 5.1-2. An example of placement of a reconstruction level which corresponds to maximum likelihood detection.

This algorithm was applied to degraded images and some experimental results are presented at the end of this chapter.

5.2 Bayesian Detector

In the last section we discussed an algorithm which performs a form of suboptimal maximum likelihood detection using a quantizer with nonuniformly spaced levels. In this section we start with the same nonuniformly spaced levels except we now develop a more flexible Bayesian detector to replace the quantizer.

To begin, briefly reivew M-hypothesis Bayesian detection from classical statistics. A detection model is usually composed of the four separate spaces as shown in Fig. 5.2-1.

I) Space $\mathcal{D}$. This is the set of all possible signals, and may be continuous or discrete. Elements of this space are denoted by D. The a priori probability distribution over the signal space is given by $f_{\mathcal{D}}(D)$, so that

$$\int_{\mathcal{D}} f_{\mathcal{D}}(D)dD = 1 \quad . \tag{5.2-1}$$

II) Space V. This is the set of all observabled. In the terminology of communication theory, a transmitted signal $D \epsilon \mathcal{D}$ becomes an observation $v \epsilon V$ due to the noise. This space is described by a conditional probability density $f_V(v|D)$, which is the probability of observing v, given a particular signal D. This density also satisfies

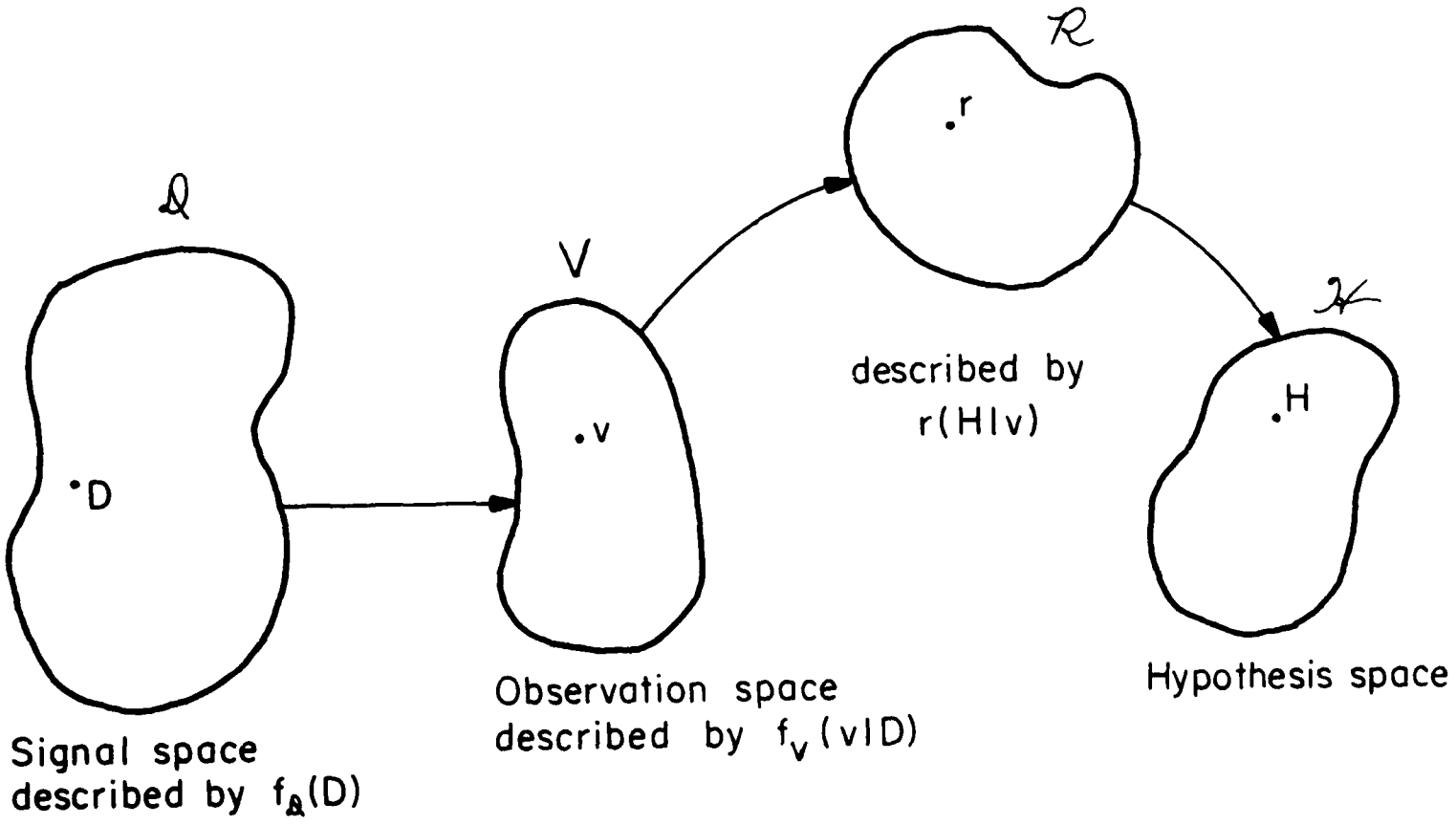


Figure 5.2-1. General detection model.

$$\int_V f_V(v|D)dv = 1. \qquad (5.2\text{-}2)$$

III) Space $\mathcal{H}$. This is the hypothesis space. Elements of $\mathcal{H}$ are various hypotheses regarding which signal D was transmitted. A particular observation v leads to the selection of an $H \in \mathcal{H}$.

IV) Space $\mathcal{R}$. This is the action space. In general we may observe a v and instead of deciding on a particular hypothesis $H_i \in \mathcal{H}$, we may decide on several hypotheses but with a probability distribution assigned to them. Elements of the space $\mathcal{R}$ are probability distribution functions and are denoted by $r(H|v)$. They denote the probability distribution over the space $\mathcal{H}$ when v is observed. Elements of $\mathcal{R}$ are called rules.

In addition to the above four spaces, there are usually three other notions that complete the detection model. They are:

i) Cost. This is denoted by C(D, H). It is a real valued function defined on space $\mathcal{D} \times \mathcal{H}$ which gives the cost of deciding on hypothesis H when signal D was transmitted.

ii) Risk. This is denoted by R(D, r) and is defined by

$$R(D, r) \triangleq \int_{\mathcal{H}} \int_V C(D, H) f_V(v|D) r(H|v) dH dv. \qquad (5.2\text{-}3)$$

This expression describes the average cost when a particular signal D is transmitted and we use rule r to decide on our hypothesis.

iii) Average risk. This is described by

$$\overline{R}(r) \triangleq \int_{\mathcal{D}} R(D, r) f_{\mathcal{D}}(D) dD \qquad (5.2\text{-}4)$$

and denotes the average risk associated with selecting a particular rule r. The averaging is performed over the signal space.

The Bayesian detector is a particular detector that observes v and decides on a rule r such that the average risk $\overline{R}(r)$ is minimized.

A special case of this model is when we limit the hypothesis space $\mathcal{H}$ to only M elements $H_1, \ldots, H_M$. In this case we divide the signal space into M regions $R_1, \ldots, R_M$ as shown in Fig. 5.2-2. Hypothesis H_i then corresponds to the assumption that the transmitted signal is from region R_i. Denoting by $P(R_i)$ the <u>a priori</u> probability that a signal might belong to region R_i and by $f_{\mathcal{D}}(D|R_i)$ the <u>a priori</u> conditional probability of the signal within region R_i we have

$$f_{\mathcal{D}}(D) = \sum_{i=1}^{M} P(R_i) f_{\mathcal{D}}(D|R_i) \ . \qquad (5.2\text{-}5)$$

One can show that in the case the minimization of the average risk results in the following detection procedure for a Bayesian detector. We calculate

$$A_j(v) = \sum_{i=1}^{M} P(R_i) \int_{R_i} C(D, H_j) f_V(v|D) f_{\mathcal{D}}(D|R_i) dD \qquad (5.2\text{-}6)$$

$$j = 1, \ldots, M$$

then we announce hypothesis H_k as being true if

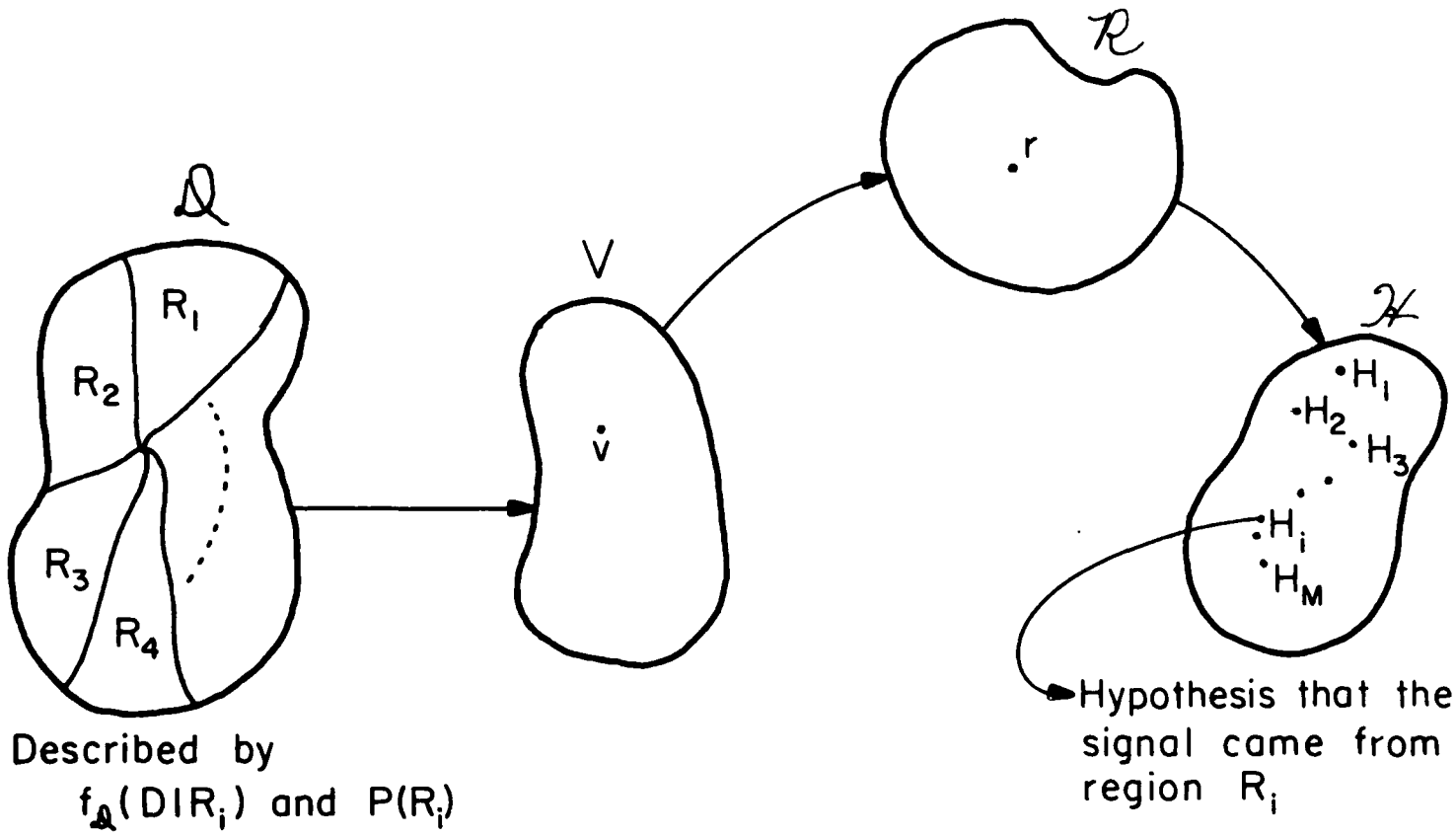


Figure 5.2-2. M hypotheses detection model.

$$A_k(v) < A_i(v) \qquad \begin{matrix} i = 1, \ldots, M \\ i \neq k \end{matrix} \quad . \tag{5.2-7}$$

Some simplifications can be made if we use the cost function

$$C(D, H_j) = \begin{cases} 0 & \text{if } D \epsilon R_j \\ 1 & \text{if } D \notin R_j \end{cases} \quad . \tag{5.2-8}$$

This is a very reasonable cost function which does not penalize correct detection. For this cost function eq. (5.2-6) reduces to

$$A_j(v) = \sum_{\substack{i=1 \\ i \neq j}}^{M} P(R_i) \int_{R_i} f_V(v|D) f_{\mathcal{D}}(D|R_i) dD \tag{5.2-9}$$

From eq. (5.2-9) it is obvious that the test rule described by eq. (5.2-7) can equivalently be stated as: for each observation v announce hypothesis H_k as being true if

$$B_k(v) > B_i(v) \qquad \begin{matrix} i = 1, \ldots, M \\ i \neq k \end{matrix} \tag{5.2-10}$$

where

$$B_i(v) = P(R_i) \int_{R_i} f_V(v|D) f_{\mathcal{D}}(D|R_i) dD \; . \tag{5.2-11}$$

With this background we now apply the Bayesian detection theory to our previous problem. Suppose we decide on M nonuniformly spaced levels in the density range of say $[0.0, 2.5]$ as discussed in the last section. These levels are selected so that the distance

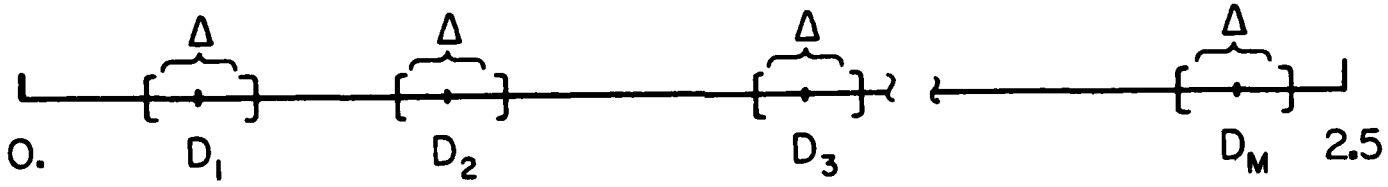
between them increases as we move up in the interval $[0.0, 2.5]$. Let us denote these levels by $D_1, \ldots, D_M$.* The detection algorithm discussed in section 5.1 involves maximum likelihood detection of regions in the degraded image having mean densities coinciding with one of these levels. The detection was implemented by using a quantizer in the first stage and followed by a second stage as discussed in section 4.4. Keeping the second stage of the detection exactly the same, we now change the first stage from maximum likelihood detection to Bayesian detection. In order to do this we start by defining regions $R_1, \ldots, R_M$ corresponding to the model of Fig. 5.2-2. Referring to Fig. 5.2-3 we let $D_1, \ldots, D_M$ be the nonuniformly spaced levels discussed earlier. We also define M intervals of width Δ

$$R_i \triangleq \left[D_i - \frac{\Delta}{2}, \; D_i + \frac{\Delta}{2} \right] \qquad i = 1, \ldots, M \tag{5.2-12}$$

where Δ is an arbitrarily small number. Now, corresponding to the detection model of Fig. 5.2-2, let the signal space be

$$\mathcal{S} = \sum_{i=1}^{M} R_i \; . \tag{5.2-13}$$

*In the last section we labeled these levels by $\ldots, Q_{-1}, Q_0, Q_1, Q_2, \ldots$ as they were also to denote quantization levels of the quantizer used in the first stage of the detection algorithm.



$$\mathcal{Q} = \sum_{i=1}^{M} R_i$$

Figure 5.2-3. The signal space used in the Bayesian detection.

In the enhancement problem the observable is the degraded image and pixels in the degraded image correspond to elements of observation space V. The enhancement procedure involves examining each pixel in the degraded image and deciding, according to eq. (5.2-10) which interval R_i it most likely belongs to. The value of the observed pixel is changed to D_i if it is decided that the pixel belongs to region R_i.

Corresponding to the signal space $\mathcal{B}$ defined inneq. (5.2-13) we can rewrite eq. (5.2-11) as

$$B_i(v) = P(R_i) \int_{D_i - \frac{\Delta}{2}}^{D_i + \frac{\Delta}{2}} f_V(v|D) f_{\mathcal{B}}(D|R_i) dD \qquad (5.2\text{-}14)$$

where $f_V(v|D)$ in the above equation is the conditional probability density of the observable v given that the signal was D. As noted in Chapter 2, when the observable is an image degraded by film-grain noise then this conditional density is Gaussian with the form

$$f_V(v|D) = \frac{1}{\sqrt{2\pi}\sigma_D} e^{-[(v-D)^2/(2\sigma_D^2)]} \qquad (5.2\text{-}15)$$

where

$$\sigma_D = k(D)^{1/3} . \qquad (5.2\text{-}16)$$

Thus if <u>a priori</u> probabilities $P(R_i)$ and $f_{\mathcal{B}}(D|R_i)$ are known, for each observed pixel value v in the degraded image, one can evaluate the following integral

$$B_i(v) = P(R_i) = \int_{D_i - \frac{\Delta}{2}}^{D_i + \frac{\Delta}{2}} \frac{1}{\sqrt{2\pi}k(D)^{1/3}} e^{-\left[(v-D)^2/(2k^2D^{2/3})\right]} \cdot f_{\mathcal{B}}(D|R_i)dD \qquad (5.2\text{-}17)$$

for $i = 1, \ldots, M$ and then according to the detection rule stated in eq. (5.2-10) change the value of the pixel to one of the M levels $D_1, \ldots, D_M$.

Since the detection procedure is to be carried out on a digital computer, evaluation of the integral of eq. (5.2-17) for each pixel of the image becomes quite time consuming. Therefore in practical use, the following simplifying assumptions are made:

1) We assume that if the signal D belongs to region R_i then it is equally likely to be anywhere in that interval. Thus we let

$$f_{\mathcal{B}}(D|R_i) = \begin{cases} \frac{1}{\Delta} & D \epsilon \left[D_i - \frac{\Delta}{2}, D_i + \frac{\Delta}{2}\right] \\ 0 & \text{elsewhere} \end{cases} . \qquad (5.2\text{-}18)$$

2) If Δ is small, one can approximate σ_D in eq. (5.2-14) by

$$\sigma_D \cong \sigma_{D_i} \triangleq k(D_i)^{1/3} \qquad (5.2\text{-}19)$$

for all D belonging to R_i. That is we neglect the signal dependence of the noise within each interval. Note, however, that we allow the variance of the noise to vary from one interval to the next.

Incorporating these two assumptions in eq. (5.2-17) we get

$$B_i(v) = \frac{P(R_i)}{\Delta} \int_{D_i - \frac{\Delta}{2}}^{D_i + \frac{\Delta}{2}} \frac{1}{\sqrt{2\pi}\, kD_i^{1/3}} \; e^{-\left[(v-D)^2\right)/(2k^2 D_i^{2/3})\right]} \, dD. \quad (5.2\text{-}20)$$

Recognizing the integral in the eq. (5.2-20) as an error function (erf), it has the closed form solution

$$B_i(v) = \frac{P(R_i)}{2\sqrt{2}\,\Delta} \left[\operatorname{erf}\left(\frac{v - D_i + \frac{\Delta}{2}}{\sigma_{D_i}} \right) - \operatorname{erf}\left(\frac{v - D_i - \frac{\Delta}{2}}{\sigma_{D_i}} \right) \right] \quad (5.2\text{-}21)$$

which reduces the evaluation of the integral to a table look up.

In summary, the Bayesian detection algorithm involves the evaluation of $B_i(v)$ for $i = 1, \ldots, M$ of eq. (5.2-21) for each pixel of the degraded image and subsequent detection according to the rule stated in eq. (5.2-10). The boundaries in the Bayesian detected image are then sharpened according to the procedure described in section 4.4.

5.3 Relation of the Bayesian Algorithm to the Noise Cheating Algorithm

In section 4.5 we noted that the "noise cheating" algorithm suggests the use of the human observer in a feedback loop to shift the quantization levels by a small amount to obtain different detected images. This would be done if the detected image had poor visual quality for a particular set of quantization levels. In fact this procedure can be repeated any number of times and the weighted sum of the various detected images can be taken to be the final enhanced image. In this section we show that under certain conditions,

performing the Bayesian algorithm once is equivalent to performing the "noise cheating" algorithm many times by shifting the quantization levels within a band of width Δ.

If a particular choice

$$f_{\mathcal{B}}(D|R_i) = \delta(D-D_i) \tag{5.2-22}$$

is made in eq. (5.2-11) we obtain

$$B_i(v) = P(R_i) f_V(v|D_i) \tag{5.2-23}$$

Furthermore, if we allow all regions in the signal space to be equally likely (i.e. $P(R_1) = P(R_2) \ldots = P(R_M)$), then the detection rule of eq. (5.2-10) reduces to deciding on hypothesis H_k if

$$f_V(v|D_k) > f_V(v|D_i) \quad \begin{matrix} i = 1, \ldots, M \\ i \neq k \end{matrix} \tag{5.2-24}$$

which is the same as the maximum likelihood detection rule of eq. (4.1-2) used in "noise cheating" algorithm. We can now observe that choosing the a priori conditional probability density function $f_{\mathcal{B}}(D|R_i)$ be a general probability density function then Bayesian algorithm is equivalent to a "noise cheating" algorithm in which we have allowed the quantization levels to vary within a band of width Δ about D_i = i = 1,...,M and where all the detected images obtained due to this shift of quantization levels are weighted by $f_{\mathcal{B}}(D|R_i)$ to obtain the final enhanced image.

5.4 Experimental Results

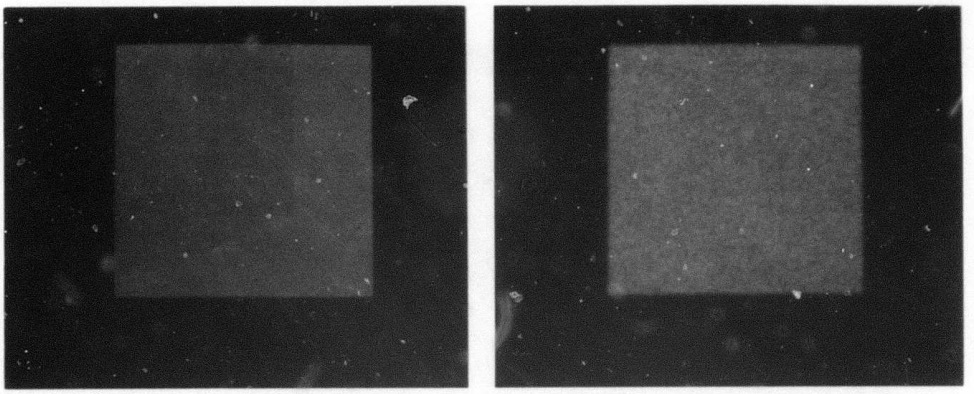
In this section we apply the different detectors discussed in this and the last chapter to some degraded images and compare their results.

Figure 5.4-1a shows a simulated block U used as the ideal image. The density levels for this image are 0.4 for the background, 1.72 for the U and 1.65 for the open slot in the U. This image of size 512 x 512 was chosen because it has a reasonably large dynamic range (0.4 to 1.72) and regions of very low contrast (the difference between the density of the U and the open slot in the U is only 0.07). In addition this image is composed of extended regions and this is of the type for which the algorithms discussed in the last two chapters are most useful.

This ideal image was degraded according to the model of Fig. 2.8-3 to obtain the image shown in Fig. 4.4-1b. Figure 5.4-1c shows the detected image using the "noise cheating" algorithm with the initial quantization level set at 0.4 and the other levels uniformly set at 4σ spacing where

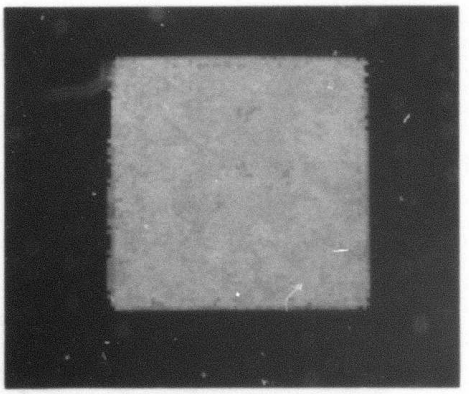
$$\sigma = k(D)^{1/3} = .3(.4)^{1/3} . \qquad (5.4\text{-}1)$$

Although some of the noise is removed by the algorithm, the result is difficult to identify. This points out a deficiency of the "noise cheating" algorithm in ignoring the signal-dependence of the



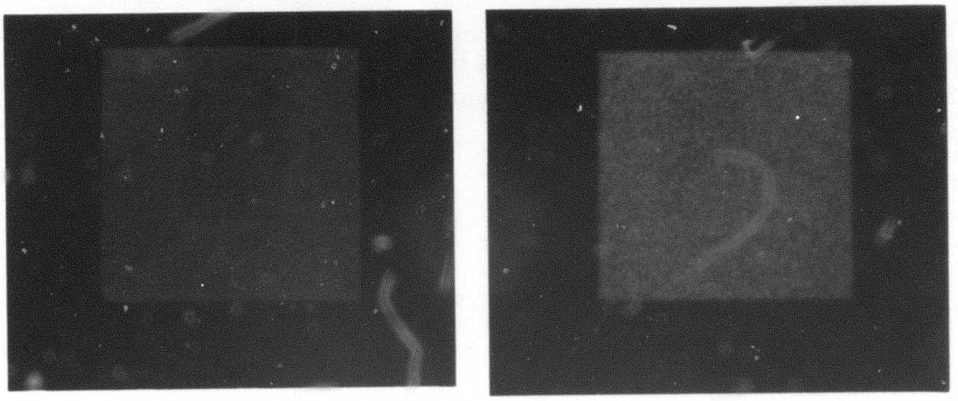
(a) Ideal image

(b) Degraded image



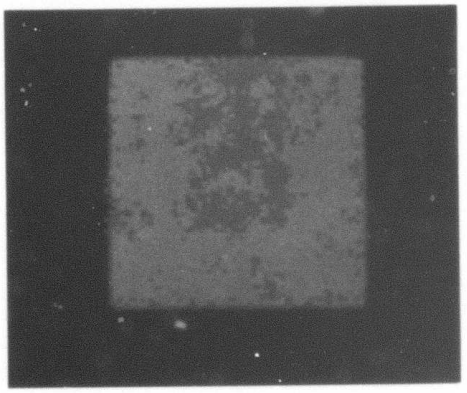
(c) Detected image using "noise cheating" algorithm

Figure 5.4-1. Application of "noise cheating" algorithm.



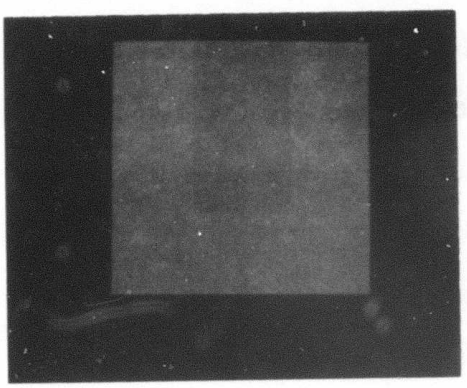
(a) Ideal image

(b) Degraded image

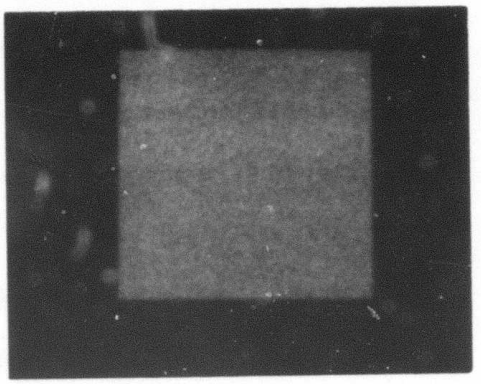


(c) Detected image

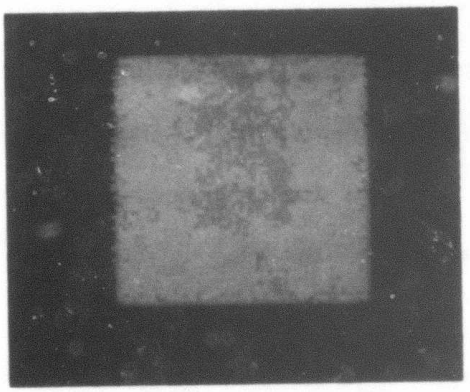
Figure 5.4-2. Application of the detection algorithm of section 5.1.



(a) Ideal image



(b) Degraded image



(d) Bayesian detected image

Figure 5.4-3. Application of Bayesian detection algorithm.

noise in enhancement of images with large dynamic range.

Figures 5.4-2c and 5.4-3c show respectively the result of applying the algorithm developed in section 5.1 and the Bayesian algorithm to the same degraded image. The block letter U is now recognizable in these pictures. Comparing Figs. 5.4-2c and 5.4-3c with Fig. 5.4-1c one sees the advantage of these algorithms over the "noise cheating" scheme.

Considering the extreme degree of the degradation, these techniques do a good job of identifying and enhancing the image features. It seems unlikely that conventional linear or nonlinear estimation techniques can produce better quality enhanced images than those shown in Fig. 5.4-2c and 5.4-3c.

We also displayed the degraded image on the and tried to adjust the brightness and the contrast knobs to see if the image could be enhanced so that the letter U would become visible on the screen. However due to extreme degree of degradation no amount of adjustment did the job.

REFERENCES

[1] M. R. Spiegle, Mathematical Handbook of Formulas and Tables, McGraw-Hill, N.Y., 1968, p. 327.

[2] D. Middleton, An Introduction to Statistical Communication Theory, McGraw-Hill, N.Y., 1960, Ch. 18.

Chapter 6

CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

This dissertation has described different means of restoring images degraded by film-grain noise. We have used some experimental and theoretical results by photographic scientist to augment existing mathematical models for imaging systems with film as the recording medium. The modeling has two primary objectives which sometimes are conflicting. The first objective has been to obtain a mathematically tractable model for subsequent estimation and detection of the degraded observed image. The second objective has been to achieve an accurate model for the physical process. A subjective test has been devised to check for the accuracy of the model. Chapter 2 of this dissertation concludes with a particular mathematical model which is a reasonable representation of various effects of imaging process without being oversimplified.

However more work in the area of modeling remains to be done. We mentioned in Chapter 2 that the standard deviation of the form given by eq. (2.3-10) is not consistant with the assumption of Gaussianity of the noisy for high and low density values. To fully understand the nature of the distributing of the noise, one needs to expose many samples of different films uniformly and digitize them

with a microdensitometer using various apertures. The success of such a procedure depends greatly on the ability to produce truly uniformly exposed samples. In past researchers doing such studies have used only 500 to 1000 points in determining the statistics of the noise. For accurate assessment many more points should be considered and statistical tests should be applied to insure that the estimates are consistent and accurate. With the improvement of microdensitometers, availability of digital computers and improvement in instrumentation, examining uniformly exposed films may finally provide some conclusive results regarding the form and the distribution of film-grain noise.

Another unexplored area in which fruitful research may be conducted is the study of the sensitivity of human vision to film-grain noise in various ranges of density. For example one can perform experiments similar to Weber's experiment [1] to determine the just noticeable contrast over a wide range of densities when the image is degraded by the signal-dependent film-grain noise. The results of such a test may be used to improve imaging system models by including the effects of the human visual system for enhancement and restoration.

In Chapter 3 we have presented a linear estimation technique which accounts for signal-dependent noise. The method is shown to be superior to conventional Wiener filtering. This area can be

further explored by considering nonlinear estimation. We have obtained the general problem of restoration of images degraded by film-grain noise in the framework of estimation theory and have shown the optimal solution to be an estimate based on nonlinear observations. The optimal estimate based on nonlinear observations is still an unsolved problem in the field of estimation theory. The most recent work in this area is due to Naraghi [2], who proposes a suboptimal recursive estimate for nonlinear observations. An interesting area for future study is the application of Naraghi's work and other nonlinear estimation techniques to images degraded by film-grain noise.

In Chapters 4 and 5 we have concentrated on the detection of a class of low contrast images degraded by severe film-grain noise. An already existing algorithm called "noise cheating" has been analyzed in the context of detection theory. Other algorithms are developed based on the same principle which are applicable to a wider class of images and are more versatile. These detectors produce fairly reasonable results when other conventional restoration methods are ineffective for extremely severe noise.

In these detection schemes each individual pixel in the degraded image is detected independent of the neighboring pixels. A future study might explore the possibility of detecting a pixel based on the result of the detection of the surrounding pixels using the correlation

among the pixels.

From the results of this study, it seems certain that future for image restoration, especially with signal-dependent noise, lies in the use of nonlinear techniques. It still remains a challenge to find efficient algorithms for rapid, high quality restoration on pictures with only a moderate amount of noise, and it is hoped that the results presented here will contribute to the development of these algorithms.

REFERENCES

[1] N. Cornsweet, Visual Perception, Academic Press, New York, 1970, p. 83.

[2] M. Naraghi, "An Algorithmic Image Estimation Method Applicable to Nonlinear Observations," University of Southern California Image Processing Institute Technical Report 580, University of Southern California, Los Angeles, California, 1975.